

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

protein - protein search, using sw model

on: June 16, 2004, 19:06:57 ; Search time 84.1687 Seconds
(without alignments)
2265.920 Million cell updates/sec

le: US-09-332-063-2

fect score: 3347

quence: 1 MRAQPSASYPVADPPA.....KTPIQILQEPDAMVEYLI 675

ring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

arched: 1586107 seqs, 282547505 residues

al number of hits satisfying chosen parameters: 1586107

imum DB seq length: 0
imum DB seq length: 2000000000

st-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

abase : A: Genesecp_29Jan04:*
1: Genesecp1980s:*
2: Genesecp1990s:*
3: Genesecp2000s:*
4: Genesecp2001s:*
5: Genesecp2002s:*
6: Genesecp2003as:*
7: Genesecp2003bs:*
8: Genesecp2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

sult No.	Score	Query Match	Length	DB ID	Description
1	3347	100.0	675	3	AAY54052 An angiot
2	3331	99.5	993	7	Aae37918 Human CGD
3	3325	99.3	675	3	AAY54053 A variant
4	2381	71.1	479	5	Abp43965 Unidentif
5	1448	43.3	637	5	Abp69085 Human pol
6	1447	43.2	608	4	Aau69439 Human Eur
7	1441	43.1	801	7	Adc38518 Human AML
8	1441	43.1	869	7	Adc37654 Human ang
9	1400	41.8	882	7	Abf56262 Mouse Jun
10	1090.5	32.6	772	7	Abf56263 Mouse Jun
11	1040.5	31.1	849	3	AAB41861 Human ORF
12	1037	31.0	467	5	Abg97502 Human NOV
13	987.5	29.5	706	4	Aab92866 Human pro
14	903	27.0	759	4	Abg17959 Novel hum
15	765.5	22.9	562	7	Adc37657 Human ang
16	712	21.3	340	4	AAM93770 Human pol
17	663	19.8	143	3	AAY54054 Angiostat
18	521.5	15.6	277	6	Abu70751 Human adi
19	328.5	9.8	182	4	AAB94793 Human pro
20	308	9.2	1072	5	Abb57023 Mouse iac
21	302	9.0	1444	4	Abg15667 Novel hum
22	296.5	8.9	1013	4	Abb71039 Drosophil
23	286	8.5	1071	4	Abb60579 Drosophil
24	285.5	8.5	566	4	Abb61040 Drosophil
25	284.5	8.5	2346	4	Abb63519 Drosophil

ALIGNMENTS

RESULT 1

RAY54052
ID AAY54052 standard; protein; 675 AA.

XX AAY54052;

XX AC

XX 27-MAR-2000 (first entry)

XX DE An angiogenesis-associated protein which binds plasminogen.

XX KW Human; angiogenesis-associated protein; plasminogen; ABP-1;

XX KW kringle domain; angiotatin; plasminogen receptor;

XX KW angiogenesis-related disease; tumor; diabetes; rheumatoid arthritis;

XX KW inflammatory disease; psoriasis; chronic inflammation; intestine; asthma;

XX KW obesity; gene therapy.

XX OS Homo sapiens.

XX FN WO9966038-A1.

XX PD 23-DEC-1999.

XX PF 11-JUN-1999; 99WO-EF004109.

XX PR 15-JUN-1998; 98SE-00002130.

XX PR 15-JUN-1998; 98US-0089266P.

XX PR 17-DEC-1998; 98SS-00004372.

XX PR 29-DEC-1998; 98US-0114386P.

XX PA (PHAA) PHARMACIA & UPJOHN AB.

XX PI Holmgren L, Troyanovsky B;

XX DR WPI; 2000-106099/09.

XX DR N-PSDB; AAZ45329.

XX PT Novel human protein useful for treating angiogenesis associated diseases

XX PT or disorders.

XX Claim 4; Page 43-46; 58pp; English.

XX The present sequence represents a human angiogenesis-associated protein

XX which is able to bind an N-terminal fragment of plasminogen. The protein

XX is designated ABP-1, and binds the first 4 kringle domains (K1-K4) and/or

XX kringle 5 (K5) of plasminogen. These four kringle domains comprise

XX angiotatin. The protein acts as a receptor for plasminogen. The

XX angiotatin-binding domain of the ABP-1 protein is described in AAY54054.

XX A polymorphic variant of ABP-1 is also described, in AAY54053. ABP-1 can

be used to manufacture medicaments for treating angiogenesis-related diseases or disorders, such as tumor conditions, diabetes, rheumatoid arthritis, and even some inflammatory diseases such as psoriasis, chronic inflammation of the intestine, asthma, etc. The protein may also be able to treat and cure, or prevent, obesity. The ABP-1 DNA can be used in gene therapy techniques

Sequence 675 AA;

Query Match 100.0%; Score 3347; DB 3; Length 675;
 Best Local Similarity 100.0%; Pred. No. 1.2e-201;
 Matches 675; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 MPRAQPSASQYQVPADPPFAIVSRAQQMVEILSDENRNLROELGECYKVARLQKVETEI 60
 1 MPRAQPSASQYQVPADPPFAIVSRAQQMVEILSDENRNLROELGECYKVARLQKVETEI 60

61 QRYSEAYENLVKSSSKREALEKAMRNKLEGEIRRMDFNDRLETRLETANKOLAEKVEYG 120
 61 QRYSEAYENLVKSSSKREALEKAMRNKLEGEIRRMDFNDRLETRLETANKOLAEKVEYG 120

121 SEDTRKTIISOLFANKKESQREKELEAEALATASSTNEDQRHIEIRDOALSNAQAKVKVL 180
 121 SEDTRKTIISOLFANKKESQREKELEAEALATASSTNEDQRHIEIRDOALSNAQAKVKVL 180

181 EELKKQVYVVKVKKMOQALVQLQACREKLEHRLTRLEERLESRIQORGNQCP 240
 181 EELKKQVYVVKVKKMOQALVQLQACREKLEHRLTRLEERLESRIQORGNQCP 240

241 TNVSEYNAALMELLREKEERILALEADMTKQKYLEENVRHFDALDAATVAAGQRTT 300
 241 TNVSEYNAALMELLREKEERILALEADMTKQKYLEENVRHFDALDAATVAAGQRTT 300

301 VISHSPNTSYDTALEARIQKEEELMANKRCCLDMEGRIKTLHAQIIKDMIKVLOQRS 360
 301 VISHSPNTSYDTALEARIQKEEELMANKRCCLDMEGRIKTLHAQIIKDMIKVLOQRS 360

361 RKEPSKTEQLSCMRPAKSLMSISNAGSGLLSHSSTLTGSPIMEEKRDDKSWKSGILIG 420
 361 RKEPSKTEQLSCMRPAKSLMSISNAGSGLLSHSSTLTGSPIMEEKRDDKSWKSGILIG 420

421 GDYRAEVVPTSPBPVPTSLLSAHSKTSRDCSTQTERGTSNKTAAVAPISVPAPVA 480
 421 GDYRAEVVPTSPBPVPTSLLSAHSKTSRDCSTQTERGTSNKTAAVAPISVPAPVA 480

481 AATAAATATATAATITTTWAAAAPVAVAAAAAATAAATAAATAAATAAATAAATAAATAA 540
 481 AATAAATATATAATITTTWAAAAPVAVAAAAAATAAATAAATAAATAAATAAATAAATAA 540

541 AASVASAAAAPASAAAAAQAQVAPAPVAPAPALVVPAPAPALVVPAPAPALVVPAPAPALV 600
 541 AASVASAAAAPASAAAAAQAQVAPAPVAPAPALVVPAPAPALVVPAPAPALVVPAPAPALV 600

601 VAPTEPPTTPPAQAQVAPASPATGPGPHRLSPSLTCNPKDGTGVPVHNTLERTPIQ 660
 601 VAPTEPPTTPPAQAQVAPASPATGPGPHRLSPSLTCNPKDGTGVPVHNTLERTPIQ 660

661 ILGQEPDAEMVEYLI 675
 661 ILGQEPDAEMVEYLI 675

SULT 2
 E37918
 AAE37918 standard; protein; 993 AA.

AAE37918;

06-NOV-2003 (first entry)

Human CGDD-7 protein.

Human; cell growth, differentiation and death protein; CGDD; leukaemia;

neurodegenerative disorder; Parkinson's disease; Alzheimer's disease; muscular disorder; myotonic dystrophy; catatonia; endocrine disorder; diabetes; Grave's disease; cancer; immunological disorder; scleroderma; systemic lupus erythematosus; allergy; Crohn's disease; renal disorder; gastrointestinal disorder; Goodpasture's syndrome; infection; cirrhosis; cardiovascular disorder; atherosclerosis; hepatic disease; transgenic; transgenic animal; gene therapy; neuroprotective; relaxant; cytostatic; dermatological; immunosuppressive; cerebroprotective; anticonvulsant; antibacterial; antiparasitic; fungicide; virucide; uropathic; cardiant; protozoacide; nootropic.

Homo sapiens.

WO2003050523-A2.

19-JUN-2003.

04-DEC-2002; 2002WO-US039133.

07-DEC-2001; 2001US-0340747P.

20-DEC-2001; 2001US-0342761P.

15-JAN-2002; 2002US-0349705P.

06-FEB-2002; 2002US-0354764P.

12-FEB-2002; 2002US-0356216P.

(INCY-) INCYTE GENOMICS INC.

Griffin JA, Rankumar J, Emerling BM, Kable AS, Elliott VS, Marquis JP, Baughn ME, Gorvad AE, Yue H, Lee EA, Becha SD, Tang YT, Tran UK, Swarnakar A, Lee S, Ison CH, Hafalia AJA, Tran B, Sprague WW, Lee SY, Khare R, Gandhi AS, Gietzen KJ, Bhatia U, Burdill JD, Blake JJ, Ho A, Zheng W;

WPI: 2003-532903/50.

N-PSDB; AAD57227.

New CGDD polypeptides, useful for diagnosing, preventing, and treating disorders associated with an abnormal expression or activity of CGDD, e.g. neuromuscular, immunological, cardiovascular disorders, cancer and/or infections.

Claim 1; Page 220-222; 299pp; English.

The present invention relates to novel cell growth, differentiation and death (CGDD) proteins and polynucleotides encoding them. The sequences of the invention are useful in diagnosing, preventing and treating disorders associated with an abnormal expression or activity of CGDD such as neurodegenerative disorders (e.g. Parkinson's disease, Alzheimer's disease), muscular disorders (e.g. myotonic dystrophy, catatonia), endocrine disorders (e.g. diabetes, Grave's disease), cancers (e.g. leukaemia, cervical or breast cancers), immunological disorders (e.g. scleroderma, systemic lupus erythematosus, allergies), gastrointestinal disorders (e.g. Crohn's disease), renal disorders (e.g. Goodpasture's syndrome), infections (e.g. viral, bacterial, fungal, parasitic, protozoal, helminthic), cardiovascular disorders (e.g. atherosclerosis) and hepatic diseases (e.g. cirrhosis). The polynucleotides can be used to create humanised animals or transgenic animals to model human diseases. The invention is also used in gene therapy. The present sequence is human CGDD-7 protein

Sequence 993 AA;

Query Match 99.5%; Score 3331; DB 7; Length 993;
 Best Local Similarity 99.8%; Pred. No. 2e-200; 2; Indels 0; Gaps 0;
 Matches 672; Conservative 1; Mismatches 1

Oy 1 MPRAQPSASQYQVPADPPFAIVSRAQQMVEILSDENRNLROELGECYKVARLQKVETEI 60
 319 MPRAQPSASQYQVPADPPFAIVSRAQQMVEILSDENRNLROELGECYKVARLQKVETEI 378

Oy 61 QRYSEAYENLVKSSSKREALEKAMRNKLEGEIRRMDFNDRLETRLETANKOLAEKVEYG 120
 379 QRYSEAYENLVKSSSKREALEKAMRNKLEGEIRRMDFNDRLETRLETANKOLAEKVEYG 438

541 AASVASAAAVAPSAASAAAVQVAPAPAPVPAPALVPVPAAPAAQASAPAQTAQPTSA 600
601 VAPTPAPTPPAVAQAEVPASPATGPHRLSTPSLTCTKPDKTGDFVHSNLTERTPIQ 660
601 VAPTPAPTPPAVAQAEVPASPATGPHRLSTPSLTCTKPDKTGDFVHSNLTERTPIQ 660

661 ILGQEPDAEMVEYLI 675
661 ILGQEPDAEMVEYLI 675

SULT 4
P43965
ABP43965 standard; protein; 479 AA.

ABP43965;

26-FEB-2003 (first entry)

Unidentified protein sequence.

Neuroprotective; immunomodulator; cancer; chromosome X; cytostatic;
anti-inflammatory; gene therapy; nutritional supplement; wound; burn;
ulcer; Alzheimer's disease; Huntington's disease;
amyotrophic lateral sclerosis; autoimmune disorder; inflammation;
vulnary.

Homo sapiens.

WO200231111-A2.

18-APR-2002.

11-OCT-2001; 2001WO-US027760.

12-OCT-2000; 2000US-00687527.

(HYSE-) HYSEQ INC.

Tang YT, Liu C, Zhou P, Asundi V, Zhang J, Zhao QA, Ren F;
Xue AJ, Yang Y, Wehrman T, Drmanac R;

WPI; 2002-426278/45.

N-PSDB; ABQ61209.

New polypeptides and their encoded proteins, useful as nutritional
sources or supplements, or in gene therapy, particularly for treating
wounds, Alzheimer's disease, amyotrophic lateral sclerosis, cancer or
inflammation.

Claim 20; SEQ ID # 868; 357pp + Sequence Listing; English.

The invention relates to 446 newly isolated polynucleotide sequences. The
activity of polynucleotides of the invention may be described as,
vulnary, neuroprotective, immunomodulator, cytostatic and anti-
inflammatory. Compositions comprising nucleic acids of the invention are
useful for treating a mammalian subject, or as nutritional sources or
supplements. These are useful in gene therapy, particularly for treating
wounds, burns or ulcers, Alzheimer's disease, Huntington's disease,
amyotrophic lateral sclerosis, autoimmune disorders, cancer or
inflammation. The nucleic acids and polypeptides are also useful in
diagnostic and research methods. The sequences given in records ABP43544-
ABP43989 represent polypeptides encoded by polynucleotides of the
invention. NOTE: The sequence data for this patent did not form part of
the printed specification, but was obtained in electronic format directly
from WIPO at ftp.wipo.int/pub/published_pct_sequences

Sequence 479 AA;

Query Match 71.1%; Score 2381; DB 5; Length 479;

Best Local Similarity 100.0%; Pred. No. 3.4e-141;

Matches 479; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 197 MQQALVOLQAACEKREQLBHLRLRLERLESLRQORQNCQPTNVSYNAAALMELLR 256
Db 1 MQQALVOLQAACEKREQLBHLRLRLERLESLRQORQNCQPTNVSYNAAALMELLR 60
QY 257 EKBERILALEADMTKWEQKYLEENVMEHFDALDAATVAAQRODTTVISHSPNTSYDTALEA 316
Db 61 EKBERILALEADMTKWEQKYLEENVMEHFDALDAATVAAQRODTTVISHSPNTSYDTALEA 120
QY 317 RIQKEEEEILMANKRCLDMEGRIKTLHAQIIIEKDAMIKVLRQSRKEPSKTBOLSCMRPA 376
Db 121 RIQKEEEEILMANKRCLDMEGRIKTLHAQIIIEKDAMIKVLRQSRKEPSKTBOLSCMRPA 180
QY 377 KSLMSISNAGSGLLSHSSLTGSPIMEEKDDKSKGSLGILLGGDYRAEYVPSPTSPVP 436
Db 181 KSLMSISNAGSGLLSHSSLTGSPIMEEKDDKSKGSLGILLGGDYRAEYVPSPTSPVP 240
QY 437 PSTPLLSAHSKTSGRDCSTQTERGTSNKTAAPISVPAPVAAATAATAATATATIT 496
Db 241 PSTPLLSAHSKTSGRDCSTQTERGTSNKTAAPISVPAPVAAATAATAATATATIT 300
QY 497 TWVAAAPVAAAAAASPATAATAATAAVSAPAAAGQIPAAASVASAAVAPSA 556
Db 301 TWVAAAPVAAAAAASPATAATAATAAVSAPAAAGQIPAAASVASAAVAPSA 360
QY 557 AAQVAVAPAAAPVPAPALVPVPAAPAAQASAPAQTAQPTSAVAPTPTPAVAQA 616
Db 361 AAQVAVAPAAAPVPAPALVPVPAAPAAQASAPAQTAQPTSAVAPTPTPAVAQA 420
QY 617 EVPASPATGPHRLSTPSLTCTKPDKTGDFVHSNLTERTPIQILGQEPDAEMVEYLI 675
Db 421 EVPASPATGPHRLSTPSLTCTKPDKTGDFVHSNLTERTPIQILGQEPDAEMVEYLI 479

RESULT 5

ABP69085

ID ABP69085 standard; protein; 637 AA.

AC ABP69085;

DT 20-JAN-2003 (first entry)

DE Human polypeptide SEQ ID NO 1132.

XX Human; genome mapping; gene therapy; food supplement; virus; fungus;
cell-proliferative disorder; neurodegenerative disease; bacterial;
Parkinson's disease; Alzheimer's disease; autoimmune disease;
multiple sclerosis; diabetes; genetic disorder; wound; burn; infection;
arthritis; cytostatic; immunomodulator; neurotropic; dermatological;
antiparkinsonian; antidiabetic; immunosuppressive; antimicrobial;
haemostatic; vulnary; fungicide; antiviral; virucide; protozoacide;
antiarthritic.

OS Homo sapiens.

PN WO200270539-A2.

XX 12-SEP-2002.

PF 05-MAR-2002; 2002WO-US005095.

PR 05-MAR-2001; 2001US-00799451.

PA (HYSE-) HYSEQ INC.

XX Tang YT, Zhou P, Goodrich RW, Asundi V, Zhang J, Zhao QA, Ren F;
PI Xue AJ, Yang Y, Ma Y, Yamazaki Y, Chen R, Wang Z, Ghosh M;
PI Wehrman T, Wang J, Wang D, Drmanac R;

XX WPI; 2002-759812/82.

DR N-PSDB; ABZ11302.

PT New polynucleotides comprising sequences assembled from expressed

122 EDTRKTSOLFPAKNKESOREKELEAEELATATARNSTNEDORRHIEIRDOALSNAQAKVVKLE 181
 130 ED-KAAEGHYASQNKFELEKELEMEELAAVRTASEDHRRHIELDOALSNAQARVVKLE 248
 182 BELKKQVYVDKVMQOALVOLQAAACEKEQLEHRLRTRLEHLSRLRTOQCGNCQPT 241
 249 BELREKQAYVEKVKLQOALTQLSACEKEQERLRTRLEHLSRLRTOQCGNGQPA 308
 242 NVSEYNAALMELLREKEERILALEADMTKWEQKYLEENVMRHFALDAATVAARDTTV 301
 309 NMPEYNAPALLELVREKEERILALEADMTKWEQKYLEESTIRHFANNAATAAARDTTI 368
 302 ISHSPNTSY-DTALREARIQKEEBEILMANRCLDMEGRIKTLHAQIIEKDAMIKVLOORS 360
 369 INHSENGSYGSSLEAHIMWEEBEEVVOANRRCCDMYTIKNLHAKIIEKDAMIKVLOORS 428
 361 RKEPSKTEQLSCMRPAKSLMSISNAGSLLSHSTLTGSPIMEEKDDKSKWKSGLIGLG 420
 429 RKDAGKTDS-SSLRPARSVPSIA-AATGTHSRQTSLSLSSQLAEBKEEKTWKGSLIGLG 486
 421 GDYRAEYVPSTPSPVPPSTPL-----SAHSKTSGRDCSTOTERGTE 462
 487 KEHHEH--ASAPLLPPPTSALSSTIASTAASSAHAKTSGKSDSTQTDKSAE 536
 488 KEHHEH--ASAPLLPPPTSALSSTIASTAASSAHAKTSGKSDSTQTDKSAE 536

SULT 7
 C38518
 ADC38518 standard; protein; 801 AA.

ADC38518;
 18-DEC-2003 (first entry)
 Human AMLP1a myosin-tail motif amino acid sequence SEQ ID NO:867.
 human; angiometin-like protein 1; AMLP1; cytostatic; gene therapy.
 Homo sapiens.
 WO2003037931-A2.
 08-MAY-2003.
 01-NOV-2002; 2002WO-US035129.
 01-NOV-2001; 2001US-0334773P.
 (AMSH) AMERSHAM BIOSCIENCES SV CORP.
 Shannon M, Phan T;
 WPI; 2003-430501/40.
 New isolated nucleic acid molecule encoding a human angiometin-like protein, useful for treating or preventing a disorder associated with decreased or increased expression or activity of AMLP1.
 Example 2; SEQ ID NO 867; 172pp; English.
 The present invention describes the human angiometin-like protein 1 (AMLP1). human AMLP1 has cytostatic activity, and can be used in gene therapy. The AMLP1 protein, nucleic acid molecules, antibodies, and compositions of the present invention can be used for treating or preventing a disorder associated with decreased or increased expression or activity of AMLP1. The present sequence represents a myosin-tail motif amino acid sequence, which is used in an example from the present invention.

Query Match 43.1%; Score 1441; DB 7; Length 801;
 Best Local Similarity 61.8%; Pred. No. 6.9e-82;
 Matches 291; Conservative 83; Mismatches 83; Indels 14; Gaps 6;

QY 2 PRAOPSSASYPVPADPAPVSRQAQWVEILSDENRNLROELGECYKVARLOKVEIQ 61
 DB 331 FQPPPAASPSQGLGDPALIVRAQWVEILTEENRVLHOELQSYDNADKLHFKFELQ 330
 QY 62 RVSEAYENLVKSSSRKHALEKAMENKLEGIIRMGDFNDRLETLERLETANKQLAEKYEYS 121
 DB 391 RISEAYESLVKSTTKRESLDKAMENKLEGIIRLHDFNDRLETLERLETANKQLAEKYEYS 450
 QY 122 EDTRKTSOLFPAKNKESOREKELEAEELATATARNSTNEDORRHIEIRDOALSNAQAKVVKLE 181
 DB 451 GD-KAAEGHYASQNKFELEKELEMEELAAVRTASEDHRRHIELDOALSNAQARVVKLE 509
 QY 182 EELKKQVYVDKVMQOALVOLQAAACEKEQLEHRLRTRLEHLSRLRTOQCGNCQPT 241
 DB 510 EELREKQAYVEKVKLQOALTQLSACEKEQERLRTRLEHLSRLRTOQCGNGQPA 569
 QY 242 NVSEYNAALMELLREKEERILALEADMTKWEQKYLEENVMRHFALDAATVAARDTTV 301
 DB 570 NMPEYNAPALLELVREKEERILALEADMTKWEQKYLEESTIRHFANNAATAAARDTTI 629
 QY 302 ISHSPNTSY-DTALREARIQKEEBEILMANRCLDMEGRIKTLHAQIIEKDAMIKVLOORS 360
 DB 630 INHSENGSYGSSLEAHIMWEEBEEVVOANRRCCDMYTIKNLHAKIIEKDAMIKVLOORS 689
 QY 361 RKEPSKTEQLSCMRPAKSLMSISNAGSLLSHSTLTGSPIMEEKDDKSKWKSGLIGLG 420
 DB 690 RKDAGKTDS-SSLRPARSVPSIA-AATGTHSRQTSLSLSSQLAEBKEEKTWKGSLIGLG 747
 QY 421 GDYRAEYVPSTPSPVPPSTPL-----LSAHSKTSGRDCSTOTERGTE 462
 DB 748 KEHHEH--ASAPLLPPPTSALSSTIASTAASSAHAKTSGKSDSTQTDKSAE 797

RESULT 8
 ADC37654
 ID ADC37654 standard; protein; 869 AA.
 AC ADC37654;
 DT 18-DEC-2003 (first entry)
 DE Human angiometin-like protein 1a SEQ ID NO:3.
 KW human; angiometin-like protein 1; AMLP1; cytostatic; gene therapy;
 QY AMLP1a.
 OS Homo sapiens.
 PN WO2003037931-A2.
 PD 08-MAY-2003.
 PF 01-NOV-2002; 2002WO-US035129.
 PR 01-NOV-2001; 2001US-0334773P.
 PA (AMSH) AMERSHAM BIOSCIENCES SV CORP.
 PI Shannon M, Phan T;
 DR WPI; 2003-430501/40.
 DR N-PSDB; ADC38522; ADC38523.
 XX New isolated nucleic acid molecule encoding a human angiometin-like protein, useful for treating or preventing a disorder associated with decreased or increased expression or activity of AMLP1.
 PS Claim 14; SEQ ID NO 3; 172pp; English.
 CC The present invention describes the human angiometin-like protein 1 (AMLP1). human AMLP1 has cytostatic activity, and can be used in gene therapy. The AMLP1 protein, nucleic acid molecules, antibodies, and

compositions of the present invention can be used for treating or preventing a disorder associated with decreased or increased expression or activity of AMLP. The present sequence represents human AMLP from the present invention.

Sequence 869 AA;

Query Match 43.1%; Score 1441; DB 7; Length 869;
Best Local Similarity 61.8%; Pred. No. 7.7e-82;
Matches 291; Conservative 83; Mismatches 83; Indels 14; Gaps 6;

2 PRAPSSASYQVPADPAIVSRQOQVILSDNRNLROELSCYKVARLOKVETIQ 61

331 PQPPAPASQQLGPDFAIVRAQOQVILTEENRVLHQELQCYDNDKLFKELQ 390

62 RVSYAYENLVKSSKRALEKAMNKLEGEIRRMHDFNRDLRERLETANKOLAEKEVGS 121

391 RISAYSELVSTTKRASLDKAMNKLEGEIRRMHDFNRDLRERLETANKOLAEKEVGS 450

122 EDTRKTSIQFARKNESQREKLEAEATARSTNEDQRRHIEIRDQALSNAQAKVKLE 191

451 GD-KAAEGHYASQNKELKEKLEMEELAAVRTASEDRHRHIBILDQALSNAQAKVKLE 509

182 EELKXKQVYDKVEMQOALVOLQACEKEOLEHRLRTLELESLRLQOQGNCOPT 241

510 EELREKQAYVEKVKQOALVOLQACEKEOLEHRLRTLELESLRLQOQGNCOPT 569

242 NVSYNAALMELREKEERILAEADMTKWEQKYLEENVMRHFALDAATAAQAQDDTTV 301

570 NMPYNAPALLEVREKEERILAEADMTKWEQKYLEENVMRHFALDAATAAQAQDDTTV 629

302 ISHSPNTSY-DTALAEARIQKEEESILMANKRCLEDMGRITKLHAQIIEKDAMIKVLOQRS 360

630 INHNRNGSYGESSLEAHIMQEEEVQANRRCCQDMETIKNLHAKIIEKDAMIKVLOQRS 689

361 RKEPSKTEQLSCMRPAKSLMSISNAGSGLSHSSTLTGSPIMEEKDDKSWKSGILIG 420

690 RXDAGKTDSS-LRPARSVPSIA-AATGTHSRQTSLSQACEKEOLEHRLRTLELESLRLQOQGNCOPT 747

421 GDYRAEVVPTSPVPSPTSL-----LSAHSKTSGRDSCSTQTERGT 462

748 KEHH-EHASAPLLPPPTSLSSIASTTAASSAHAKTSGKSDSTQTDKSAE 797

SULT 9
R56262

ABR56262 standard; protein; 882 AA.

ABR56262;

20-NOV-2003 (first entry)

Mouse Junction-Enriched and Associated Protein, JEAP.

Murine; JEAP; Junction-Enriched and Associated Protein; tight junction; exocrine gland; pancreas; submandibular gland; lacrimal gland; parotid gland; sublingual gland.

Mus musculus.

EPI312615-A2.

21-MAY-2003.

15-NOV-2002; 2002EP-00079841.

16-NOV-2001; 2001JP-00352241.

(EISA) EISA CO LTD.

Nishimura M, Asano M, Ono Y, Morimoto K, Takeuchi M, Inoue Y;
Imai T, Takai Y;

DR WPI: 2003-571281/54.

XX N-PSDB; ACC70742.

PT New junction-enriched and associated protein 2 which is expressed specifically at tight junctions in exocrine glands, useful for identifying substances which affect the expression, activity or distribution of the protein.

XX Claim 13; Page 15-17; 3lpp; English.

XX The present sequence is the protein sequence for murine Junction-Enriched and Associated Protein (JEAP). JEAP is expressed specifically at tight junctions in exocrine glands such as the pancreas, submandibular gland, lacrimal gland, parotid gland and sublingual gland. JEAP is useful for identifying substances which affect its expression, activity or distribution

XX Sequence 882 AA;

Query Match 41.8%; Score 1400; DB 7; Length 882;
Best Local Similarity 58.2%; Pred. No. 2.9e-79;
Matches 299; Conservative 80; Mismatches 103; Indels 32; Gaps 11;

QY 1 MPRAQ-SSASYQVPADP-----FAIVSRQOQVILSDNRNLROELSCYKVA 51

333 LPLPLPISLAASQPLPASPNQQLGPDFAIVRAQOQVILTEENRVLHQELQCYDND 392

QY 52 RLOQVETIQVSEAYENLVKSSKRALEKAMNKLEGEIRRMHDFNRDLRERLETANK 111

393 KLHKEFELQSLSEAYSELVSTTKRASLDKAMNKLEGEIRRMHDFNRDLRERLETANK 452

QY 112 QLAKYEGESDTRKTSIQFARKNESQREKLEAEATARSTNEDQRRHIEIRDQALS 171

453 QLSREYDGHED-KAABSHYVSNQKELKEKLEMEELAAVRTASEDRHRHIBILDQALS 511

QY 172 NQAQKVKLEELKQVYDKVEMQOALVOLQACEKEOLEHRLRTLELESLRLQOQGNCOPT 231

512 NQAQKVKLEELKQVYDKVEMQOALVOLQACEKEOLEHRLRTLELESLRLQOQGNCOPT 571

QY 232 QQRQGNCOPTNVSEYNAALMELREKEERILAEADMTKWEQKYLEENVMRHFALDAATA 291

572 QQKHGTGPPVPSLPECNAPALMELVREKEERILAEADMTKWEQKYLEENVMRHFALDAATA 631

QY 292 TVAAQRDTTVISHSPNTSY-DTALAEARIQKEEESILMANKRCLEDMGRITKLHAQIIEKD 350

632 AATAERDTTISNHRNGSYGESSLEAHIMQEEEVQANRRCCQDMETIKNLHAKIIEKD 691

QY 351 AMIKVLOQRSKPSKTEQLSCMRPAKSLMSISNAGSGLSHSSTLTGSPIMEEKDDK- 409

692 AMIKVLOQRSKPSKTEQLSCMRPAKSLMSISNAGSGLSHSSTLTGSPIMEEKDDK- 749

QY 410 SWKSGILIG-----GDYRAEVVPTSPVPSPTSL-----LSAHSKTSGRDSCSTQTERGT 461

750 TWKSGIFGLGKEHQOQASAPLPTTPASALSPLASTTSASSTHAKTSGKSDSTQTDKST 809

QY 462 E-----SNKTAAPVAPISVPAPVAAAATAA 485

810 ELFWPSNALSPLSGRLSTAPSNPILKHPAKGA 843

RESULT 10

ABR56263

ID ABR56263 standard; protein; 772 AA.

XX ABR56263;

AC ABR56263;

XX 20-NOV-2003 (first entry)

DE Mouse Junction-Enriched and Associated Protein-2, JEAP-2.

XX Murine; JEAP-2; Junction-Enriched and Associated Protein-2;

KW exocrine gland; pancreas; submandibular gland; lacrimal gland;

KW tight junction.

X Mus musculus.
S
N EP1312615-A2.
D 21-MAY-2003.
X F 15-NOV-2002; 2002EP-00079841.
X R 16-NOV-2001; 2001JP-00352241.
X K (EISA) EISAI CO LTD.
A Nishimura M, Asano M, Ono Y, Morimoto K, Takeuchi M, Inoue Y;
I Imai T, Takai Y;
I WPI; 2003-571281/54.
I N-PSDB; ACC70743.
I New junction-enriched and associated protein 2 which is expressed
I specifically at tight junctions in exocrine glands, useful for
I identifying substances which affect the expression, activity or
I distribution of the protein.
I Claim 1; Page 20-22; 31pp; English.
I The present sequence is the protein sequence for murine Junction-Enriched
I and Associated Protein-2 (JAP-2). JAP-2 is expressed specifically at
I tight junctions in exocrine glands such as the pancreas, submandibular
I gland and lacrimal gland. JAP-2 is useful for identifying substances
I which affect its expression, activity or distribution
I Sequence 772 AA;
Query Match 32.6%; Score 1090.5; DB 7; Length 772;
Best Local Similarity 50.4%; Pred. No. 6.6e-60;
Matches 244; Conservative 71; Mismatches 118; Indels 51; Gaps 9;
6 PSSASYQVPADPPFAIV-----SRAQQVVEILSD-----ENRNLQBLEGCGYKVARLQK 55
284 PSSFGPVAEGPSSAQATGSHLAQMETVLEARNARLQDNERLQRELESTSEKAGRIEK 343
56 VETETQVSEAYENVKSSKREALEKAMKVLGEGETRMHDFNRDLRELEFANKOLAE 115
344 LENEIQLSEAHESLMRTSSKREALEKMTNRNDGEMERLQDFNRDLRELEFANKOLAE 403
116 KEYEGSETRKTIQSLPAKKNESQREKLEAEATARSTNEQDQRIEIRDOALSNAQA 175
404 KTOEAQAGSQDWAKLLAQSYEQOQKLEKREVALLRGAIEDORRHABLELQALNAQS 463
176 KVKLEBELKXQVYVDVKKXQALVOLQACCKREQLERHLRLELESLRQORQ 235
464 RAARAEBELKXQVYVKKVLELQALQALQALQALQALQALQALQALQALQALQALQ 523
236 -----GNCQPTNVSEYNAALMELLREKBERILALEADMTKQKYLENVVHFPALDA 289
524 TGLTAGGGSGGGAELSAULSLSQLEKEHQILALEADMTKQKYLENVVHFPALDA 583
290 AATVAAQRDTTVISHSPNTSYDTALEARIQKEBEIILMANKRLCLDMEGRIKTLHAQIIEK 349
584 AATAAQRDTTLIRHSPQSPSSSF-----NEGLLPGNHRHQEMESRLKVLHAQILEK 636
350 DAMIKYLOQRKPKTEQLSCORPAKSLMSISNAGSLSSSTUTGSPINKEVRDDK 409
637 DAVIKVLQQRKPKPGKATQ-GTLRPAKSPVSPFAAAG-----TQ 676
410 SWKSGILGLGQVRAEYVPS---TPGPVPPSTPLLSAHSKTSRDCSTOTERTGTSNKT 466
677 CQWQ-----LVSSRQTDARPDGVRPAEPPATAPLPAHTKHSGRDSGTQD-GPADNTS 731
467 AAVA 470
732 ACLA 735

RESULT 11
AAB41861
ID AAB41861 standard; protein; 849 AA.
XX
AC AAB41861;
XX
DT 08-FEB-2001 (first entry)
XX
DE Human ORFX ORF1625 polypeptide sequence SEQ ID NO:3250.
XX
KW Human; open reading frame; ORFX; detection; cytostatic; hepatotropic;
KW vulnary; antipariatic; antiparkinsonian; nootropic; neuroprotective;
KW anticonvulsant; osteopathic; antiarthritic; immunosuppressant; cardiant;
KW immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic;
KW hypotensive; dermatological; immunosuppressive; antiinflammatory;
KW antiviral; antibacterial; antifungal; antirheumatic; antithyroid;
KW antianemic; gene therapy; cancer; proliferative disorder; hypertension;
KW neurodegenerative disorder; osteoarthritis; graft vs host disease;
KW cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;
KW cholesterol ester storage; systemic lupus erythematosus; infection;
KW severe combined immunodeficiency; malaria; autoimmune disorder; asthma;
KW allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;
KW bone damage; cartilage damage; antiinflammatory disease; coagulation;
KW thrombosis; contraceptive.
XX
XX Homo sapiens.
OS
XX
XX WO200058473-A2.
PN
XX
XX 05-OCT-2000.
PD
XX
XX 31-MAR-2000; 2000WO-US008621.
PF
XX
XX 31-MAR-1999; 99US-0127607P.
PR
XX 02-APR-1999; 99US-0127636P.
PR
XX 05-APR-1999; 99US-0127728P.
PR
XX 30-MAR-2000; 2000US-00540763.
PR
XX
XX (CURA-) CURAGEN CORP.
PA
XX
XX Shinketsu RA, Leach M;
PI
XX
XX WPI; 2000-602362/57.
DR
XX N-PSDB; AAC76070.
DR
XX
XX Novel nucleic acids and peptides derived from open reading frame X,
PT useful for treating e.g. cancers, proliferative disorders,
PT neurodegenerative disorders and cardiovascular disease.
XX
XX Claim 11; Page 2453-2455; 5507pp; English.
XX
XX AAC74446 to AAC77506 encode the proteins given in AAB40237 to AAB43387,
XX which represent the human ORFX open reading frames 1 to 3181. The ORFX
XX sequences have activities such as: cytostatic; hepatotropic; vulnary;
XX antipariatic; antiparkinsonian; nootropic; neuroprotective; osteopathic;
XX anticonvulsant; antiarthritic; immunosuppressant; immunostimulant;
XX cardiant; thrombolytic; coagulant; vasotropic; antidiabetic; hypotensive;
XX dermatological; immunosuppressive; antiinflammatory; antibacterial;
XX antiviral; antifungal; antirheumatic; antithyroid; and antianemic. The
XX sequences can be used for determining the presence of or predisposition
XX to, or preventing or treating pathological conditions associated with an
XX ORFX-associated disorder. The nucleic acids can be used to express ORFX
XX proteins in gene therapy vectors. The proteins and nucleic acids may be
XX used to treat cancers, proliferative disorders, neurodegenerative
XX disorders, osteoarthritis, graft vs host disease, cardiovascular disease,
XX diabetes mellitus, hypertension, hypothyroidism, cholesterol ester
XX storage, systemic lupus erythematosus, severe combined immunodeficiency
XX (SCID), AIDS, viral, bacterial or fungal infection, malaria, autoimmune
XX disorders, asthma, allergies, aplastic anaemia, burns, wounds, bone and
XX cartilage damage, nocturnal haemoglobinuria, antiinflammatory disease; to
XX enhance coagulation; to inhibit thrombosis; and as a contraceptive

Query Match 31.0%; Score 1037; DB 5; Length 467;
Best Local Similarity 48.4%; Pred. No. 8e-57;
MATCHES 233; Conservative 75; Mismatches 117; Indels 56; Gaps 9;

32 LSPDENRLQELGCVKVARLOKVETIQRVSEAYENLVKSSKEALEKAMRNKLGE 91
| | | | | : | : | : | : | : | : | : | : | : | : | : | : |
11 LQDNERLQRELESRAEKRIEKSIOQLSEHSLTRASSKEALEKTWRNMUSE 70
| | | | | : | : | : | : | : | : | : | : | : | : | : | : |
92 IRMHPNDRDLRERLETANKOLAEKEYEGSBTDRTKIISLFKNYKGSOREKKILAEUAT 151
| | | | | : | : | : | : | : | : | : | : | : | : | : | : |
71 MRRLQDPNRDLRERLESAENRRLLASKTQAAGSQDMVAKLLAQSYEQOQEKLREMAAL 130
| | | | | : | : | : | : | : | : | : | : | : | : | : | : |
152 ARSTNEQORRHIEIRDOALSNAQAKVKVLEELKKOVVDYKVMWOOALVOLQAACEKR 211
| | | | | : | : | : | : | : | : | : | : | : | : | : | : |
131 ERGAIEDQRRRELLEQALGNAGQAPAAEEELRKQAIYEKVERIQQAQGLOLAACEKR 190
| | | | | : | : | : | : | : | : | : | : | : | : | : | : |
212 EOLEHLARTLRERELESRLIQRO-----GCNPNTNVSYNNAAAAMELLREKEERILALE 266
| | | | | : | : | : | : | : | : | : | : | : | : | : | : |
191 EQELELRTRLBOELKAURLAQQRAQAGPGGSGSGGFELSARLSGLREKEEQI LALE 250
| | | | | : | : | : | : | : | : | : | : | : | : | : | : |
267 ADMTKWEOKVLBNVRHFALDAATVAARDTTTVISHSENTSYDTALBARIQKEEBEIL 326
| | | | | : | : | : | : | : | : | : | : | : | : | : | : |
251 ADMTKWEOKYLERRAVRFAMDAAATAAAADTTTTLIHSFPQPSPSSF-----NBGLL 303
| | | | | : | : | : | : | : | : | : | : | : | : | : | : |
327 MANKRCLDEMGRIKTLHAQIIIEKDAMI KVALQORKEPKSKTEQLSCMRPAKSLMISTSNAG 386
| | | | | : | : | : | : | : | : | : | : | : | : | : | : |
304 TCGHRHQEMESRLKVLHAQILEKDAVIKVALQRSERPFGKAI QGS-LRPAKSVPSVFANA 362
| | | | | : | : | : | : | : | : | : | : | : | : | : | : |
387 SGLLSHSSLTIGSPIWEEKRDDKSWKSGLGILLGDYZRAYVPVS-----TPSPVPP 437
| | | | | : | : | : | : | : | : | : | : | : | : | : | : |
363 AA-----GTQGWOG----LSSSERQTADAPARLTADRAPTEEPVVT 400
| | | | | : | : | : | : | : | : | : | : | : | : | : | : |
438 STPLLGAHKSTGRDCSTOTERTGESNKTAAVAIPISVPAP-----VAAAAATAAITATAA 492
| | | | | : | : | : | : | : | : | : | : | : | : | : | : |
401 APP--AAAHAKHGSRDGSTQTDGPPTDSTICL-----PEPDLSULGCCSQRASLDVSAT 453
| | | | | : | : | : | : | : | : | : | : | : | : | : | : |
493 T 493
:
454 S 454

SUBMIT 13
AB92866
AA92866 standard; protein; 706 AA.
AB92866;
26-JUN-2001 (first entry)
Human protein sequence SEQ ID NO:11448.
Human; primer; detection; diagnosis; antisense therapy; gene therapy.
Homo sapiens.
EF1074617-A2.
07-FEB-2001.
28-JUL-2000; 2000EP-00116126.
29-JUL-1999; 99JP-00248036.
27-AUG-1999; 99JP-00300253.
11-JAN-2000; 2000JP-00118776.
02-MAY-2000; 2000JP-00183767.
09-JUN-2000; 2000JP-00241899.

(HELI-) HELIX RES INST.

Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
WPI; 2001-318749/34.

```

XX Primer sets for synthesizing polynucleotides, particularly the 5602 full-
PT length cDNAs defined in the specification, and for the detection and/or
PT diagnosis of the abnormality of the proteins encoded by the full-length
PT cDNAs.
XX
PS Claim 8; SEQ ID NO 11448; 2537bp + Sequence Listing; English.
XX
CC The present invention describes primer sets for synthesizing 5602 full-
CC length cDNAs defined in the specification. Where a primer set comprises:
CC (a) an oligo-dr primer and an oligonucleotide complementary to the
CC complementary strand of a polynucleotide which comprises one of the 5602
CC nucleotide sequences defined in the specification, where the
CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination
CC of an oligonucleotide comprising a sequence complementary to the
CC complementary strand of a polynucleotide which comprises a 5'-end
CC sequence and an oligonucleotide comprising a sequence complementary to a
CC polynucleotide which comprises a 3'-end sequence, where the
CC oligonucleotide comprises at least 15 nucleotides and the combination of
CC the 5'-end sequence/3'-end sequence is selected from those defined in the
CC specification. The primer sets can be used in antisense therapy and in
CC gene therapy. The primers are useful for synthesizing polynucleotides,
CC particularly full-length cDNAs. The primers are also useful for the
CC detection and/or diagnosis of the abnormality of the proteins encoded by
CC the full-length cDNAs. The primers allow obtaining of the full-length
CC cDNAs easily without any specialised methods. AA03166 to AAH13628 and
CC AAH13633 to AAH18742 represent human cDNA sequences; AA092446 to AA095993
CC represent human amino acid sequences; and AAH13629 to AAH13632 represent
CC oligonucleotides, all of which are used in the exemplification of the
CC present invention
XX
CC Sequence 706 AA;
XX
QY Query Match 29.5%; Score 987.5; DB 4; Length 706;
XX Best Local Similarity 52.3%; Pred. No. 1.7e-53;
XX Matches 214; Conservative 65; Mismatches 101; Indels 29; Gaps 5;
XX
QY 13 PYPADPPAIVSRAQOMVEI-----LSDENRNLROBLEGCYKVARLQKVETIQRV 63
XX ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
DB 296 PVSQAASATSGSAHLAQMBAVLFNARLQDRNERLQRELESSAERKAEIKLESEIQR 355
XX ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
QY 64 SEAYENLVKSSKREALEKAMRNKLEGEIRRMHDFNRDLRERLETANKOLAEKEYGSD 123
XX ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
DB 356 SEAHSLTRASKKREALEKTRNNKMDSEMRRLQDFRDLRERLESANRLASTQEAQAG 415
XX ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
QY 124 TRKTIISQLPAXNKSESRFEKGLAEALATARNSTVEDRRHIEIRDOALSNAQKVKLEEE 183
XX ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
DB 416 SDQMVKLLAQSYEQEQEQLEREMALLGAIEDQRRRAELLEQALGNAQGRAARABEE 475
XX ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
QY 184 LKKKQVYVYKVKQOALVOLQACERQELERLRLTRLEERLESRLRIQRO-----GNC 238
XX ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
DB 476 LRKKQIVYKVKERLQOALGQOACERQELERLRLTRLEQELKALRAQORQAGAPGSS 535
XX ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
QY 239 QPTNVSYNAALMELRREKEERILALEADMTKWQKYLEENVRHFALDAATAATVAAQRD 298
XX ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
DB 536 GSGSGPFLSRLRSEQLRKEBEQILALEADMTWQKYLEEHAMRFQAMDAATAATAAQRD 595
XX ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
QY 299 TTWISHPNTSVDTALEARLQKBEETILMANKCLDWBGRITKLHAIITEKDMATKVLQ 358
XX ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
DB 596 TTLIRHSFQFPSSFP-----NEGTLTGHRHQHMESELKVLVHAIILEKDAVILVQ 648
XX ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
QY 359 RGRKFSKTEQLSCHMRPAKSLMGSISNAGS-----GLLSHSTLTGSP 400
XX ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
DB 649 RSRDPGRKAIQGS--LRPAKSPVSVFAAAAAAGTQGWGLSSSRQTADAP 696
XX ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
XX
RESULT 14
ABGI7959
ID ABGI7959 standard; protein; 759 AA.
XX
AC ABGI7959;
XX
TT 18-FEB-2002 (first entry)

```

RESULT 14	
ABG17959	
ID ABG1	
XX	
AC ABG1	
XX	
DT 18-F	

Db 705 EEV 708

RESULT 15

ADC37657

ID ADC37657 standard; protein; 562 AA.

XX

AC ADC37657;

XX

DT 18-DEC-2003 (first entry)

XX

DE Human angiotensin-like protein 1b SEQ ID NO:6.

XX

XX human; angiotensin-like protein 1; AMLP1; cytostatic; gene therapy;

KW AMLP1b.

KW

XX Homo sapiens.

OS

XX WO20003037931-A2.

PN

XX 08-MAY-2003.

PD

XX

PF 01-NOV-2002; 2002WO-US035129.

XX

XX 01-NOV-2001; 2001US-0334773P.

PR

XX (AMSH) AMERSHAM BIOSCIENCES SV CORP.

PA

XX Shannon M, Phan T;

PI

XX WPI; 2003-430501/40.

DR

DR N-PSDB; ADC37655, ADC37656.

XX

XX New isolated nucleic acid molecule encoding a human angiotensin-like

PT protein, useful for treating or preventing a disorder associated with

PT decreased or increased expression or activity of AMLP1.

XX

PS Claim 14; SEQ ID NO 6; 172pp; English.

XX

XX The present invention describes the human angiotensin-like protein 1

CC (AMLP1). human AMLP1 has cytostatic activity, and can be used in gene

CC therapy. The AMLP1 protein, nucleic acid molecules, antibodies, and

CC compositions of the present invention can be used for treating or

CC preventing a disorder associated with decreased or increased expression

CC or activity of AMLP1. The present sequence represents human AMLP1b from

CC the present invention.

XX

SQ Sequence 562 AA;

Query Match 22.9%; Score 765.5; DB 7; Length 562;

Best Local Similarity 65.5%; Pred. No. 1.1e-39;

Matches 152; Conservative 39; Mismatches 40; Indels 1; Gaps 1;

QY 2 PRAPSSASVQVPADPPFAIVSRAQOMVEILSDENRNLROELGCGYKVARLOKVEIEIQ 61

DB 332 PQQPAPASVQQLGPDFAIVRAQOMVEILSDENRNLROELGCGYKVARLOKVEIEIQ 391

QY 62 RVSEAYENLVKSSSKREALEKAMRNKLEGEIRRMHDFNRDLERLETANKQLAEKVEYS 121

DB 392 RISEAYESLVKSTTKRESLDKAMRNKLEGEIRRLHDFNRDLERLETANKQLAEKVEYS 451

QY 122 EDTKRTISQLFANKKESQREKLEAELETARSTNEDQRRHIEIRDOALSNAQAKVKLE 181

DB 452 ED-KAAEGHYASQNKFFLEKEBELEMELEAAVATASDEDRHIEIRDOALSNAQAKVKLE 510

QY 182 EELKKKQVYVDKVKEMQOALVQLQAACEKLEQLERLRLERLESLRQQ 233

DB 511 EELREKAYVEKVEKLEKQALVQLQAACEKLEQLERLRLERLESLRQQ 562

Search completed: June 16, 2004, 19:12:37

Job time : 88.1687 secs

Novel human diagnostic protein #17950.

Human; chromosome mapping; gene mapping; gene therapy; forensic; food supplement; medical imaging; diagnostic; genetic disorder.

Homo sapiens.

WO200175067-A2.

11-OCT-2001.

30-MAR-2001; 2001WO-US008631.

31-MAR-2000; 2000US-00540217.

23-AUG-2000; 2000US-00649167.

(HYSE-) HYSEQ INC.

Drmanac RT, Liu C, Tang YT;

WPI; 2001-639362/73.

N-PSDB; AAS82146.

New isolated polynucleotide and encoded polypeptides, useful in diagnostics, forensics, gene mapping, identification of mutations responsible for genetic disorders or other traits and to assess biodiversity.

Claim 20; SEQ ID NO 48318; 103pp; English.

The invention relates to isolated polynucleotide (I) and polypeptide (II) sequences. (I) is useful as hybridisation probes, polymerase chain reaction (PCR) primers, oligomers, and for chromosome and gene mapping, and in recombinant production of (II). The polynucleotides are also used in diagnostics as expressed sequence tags for identifying expressed genes. (I) is useful in gene therapy techniques to restore normal activity of (II) or to treat disease states involving (II). (II) is useful for generating antibodies against it, detecting or quantitating a polypeptide in tissue, as molecular weight markers and as a food supplement. (II) and its binding partners are useful in medical imaging of sites expressing (II). (I) and (II) are useful for treating disorders involving aberrant protein expression or biological activity. The polypeptide and polynucleotide sequences have applications in diagnostics, forensics, gene mapping, identification of mutations responsible for genetic disorders or other traits to assess biodiversity and to produce other types of data and products dependent on DNA and amino acid sequences. ABG0010-ABG30377 represent novel human diagnostic amino acid sequences of the invention. Note: The sequence data for this patent did not appear in the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published_pat_sequences

Sequence 759 AA;

Query Match 27.0%; Score 903; DB 4; Length 759;

Best Local Similarity 99.5%; Pred. No. 3.8e-48;

Matches 183; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

1 MPRAQPSASVQVPADPPFAIVSRAQOMVEILSDENRNLROELGCGYKVARLOKVEIEI 50

525 MPRAQPSASVQVPADPPFAIVSRAQOMVEILSDENRNLROELGCGYKVARLOKVEIEI 584

61 QRVSEAYENLVKSSSKREALEKAMRNKLEGEIRRMHDFNRDLERLETANKQLAEKVEYS 120

595 QRVSEAYENLVKSSSKREALEKAMRNKLEGEIRRMHDFNRDLERLETANKQLAEKVEYS 644

121 SEDTKRTISQLFANKKESQREKLEAELETARSTNEDQRRHIEIRDOALSNAQAKVKLE 180

645 SEDTKRTISQLFANKKESQREKLEAELETARSTNEDQRRHIEIRDOALSNAQAKVKLE 704

181 EEL 184

Sequence 3, Appli
Sequence 31, Appl
Sequence 54, Appl
Sequence 2, Appli
Sequence 2, Appli
Sequence 85, Appl
Sequence 12, Appl
Sequence 1, Appli
Sequence 3, Appli
Sequence 4, Appli
Sequence 4, Appli
Sequence 47, Appl
Patent No. 5427901
Patent No. 5482709
Sequence 2, Appli
Sequence 6, Appli
Sequence 6, Appli

28 209.5 6.3 1315 3 US-08-899-595-3
29 209.5 6.3 2662 4 US-09-595-684B-31
30 209 6.2 219 2 US-08-557-309B-54
31 208.5 6.2 1248 2 US-09-080-897-2
32 208.5 6.2 1248 3 US-09-323-735-2
33 207 6.2 803 4 US-09-154-750A-85
34 207 6.2 803 4 US-09-665-479A-12
35 206.5 6.2 1939 4 US-09-310-187A-1
36 205.5 6.1 1886 4 US-08-938-105-3
37 204 6.1 2101 1 US-08-195-487-4
38 204 6.1 2101 5 PCT-US93-06160-4
39 203.5 6.1 550 4 US-09-616-289-47
40 203 6.1 180 6 5273901-7
41 203 6.1 180 6 5482709-6
42 203 6.1 576 2 US-08-533-306A-2
43 203 6.1 576 2 US-08-742-923A-2
44 203 6.1 816 2 US-08-533-306A-6
45 203 6.1 816 2 US-08-742-923A-6

ALIGNMENTS

RESULT 1

US-08-714-741-41
; Sequence 41 Application US/08714741
; Patent No. 6506613
; GENERAL INFORMATION:
; APPLICANT: Briles, David E.
; APPLICANT: McDaniel, Larry S.
; APPLICANT: Swiatlo, Edwin
; APPLICANT: Yother, Janet
; APPLICANT: Crain, Marilyn J.
; APPLICANT: Hollingshead, Susan
; APPLICANT: Hart, Rebecca
; APPLICANT: Brooks-Walter, Alexis
; TITLE OF INVENTION: PNEUMOCOCCAL GENES, PORTIONS THEREOF,
; TITLE OF INVENTION: EXPRESSION PRODUCTS THEREFROM, AND USES OF SUCH GENES,
; NUMBER OF SEQUENCES: 47
; CORRESPONDENCE ADDRESS:
; ADDRESSES: Curtis, Morris & Safford, P.C.
; STREET: 530 Fifth Avenue
; CITY: New York
; STATE: New York
; COUNTRY: U.S.
; ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/714,741
; FILING DATE: 16-SEP-1996
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Frommer Esq., William S.
; REGISTRATION NUMBER: 25,506
; REFERENCE/DOCKET NUMBER: 454312-2460
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 840-3333
; TELEFAX: (212) 840-0712
; INFORMATION FOR SEQ ID NO: 41:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1231 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: amino acid
US-08-714-741-41

Query Match 8.2%; Score 275; DB 4; Length 1231;

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

protein - protein search, using sw model

on: June 16, 2004, 19:08:12 ; Search time 28.0562 Seconds
(without alignments)
1242.060 Million cell updates/sec

le: US-09-332-063-2

fect score: 3347
uence: 1 MPRAQPSASVQVFADPFA.....KTPIQIQBPAEMVEYLI 675

ring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

arched: 389414 seqs, 51625971 residues

al number of hits satisfying chosen parameters: 389414

imum DB seq length: 0
imum DB seq length: 200000000

st-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

abase : Issued Patents AA:*
1: /cgn2_6/prodata/2/iaa/5A_COMB.pep:*
2: /cgn2_6/prodata/2/iaa/5B_COMB.pep:*
3: /cgn2_6/prodata/2/iaa/6A_COMB.pep:*
4: /cgn2_6/prodata/2/iaa/6B_COMB.pep:*
5: /cgn2_6/prodata/2/iaa/POCUS_COMB.pep:*
6: /cgn2_6/prodata/2/iaa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

sult No.	Score	Query Match	Length	DB ID	Description
1	275	8.2	1231	4	US-08-714-741-41
2	263	7.9	8991	4	US-08-714-741-32
3	252	7.5	2972	4	US-09-579-181-2
4	252	7.5	3118	4	US-09-579-181-1
5	250.5	7.5	399	4	US-09-252-991A-22853
6	237.5	7.1	803	3	US-09-103-429A-4
7	235	7.0	786	3	US-09-103-429A-3
8	234.5	7.0	1326	4	US-09-688-188B-15
9	234.5	7.0	1326	4	US-09-291-417D-15
10	231	6.9	316	4	US-09-252-991A-32957
11	226.5	6.8	2482	1	US-08-328-254-6
12	225.5	6.7	756	4	US-09-963-137-184
13	225	6.7	3248	1	US-08-353-700-1
14	225	6.7	3248	5	PCT-US95-16216-1
15	223.5	6.7	1079	4	US-09-489-039A-7502
16	216	6.5	442	3	US-08-834-306-52
17	216	6.5	442	3	US-08-993-674A-52
18	216	6.5	442	4	US-09-256-976-52
19	213	6.4	262	1	US-08-403-379A-1
20	213	6.4	262	2	US-08-929-414-1
21	213	6.4	263	2	US-08-557-309B-51
22	213	6.4	263	3	US-08-834-306-51
23	213	6.4	263	3	US-08-993-674A-51
24	213	6.4	263	4	US-09-256-976-51
25	213	6.4	1972	4	US-08-875-435B-3
26	212.5	6.3	1274	4	US-09-095-443-2
27	210.5	6.3	1075	4	US-09-252-991A-18387

Best Local Similarity 22.3%; Pred. No. 4.5e-10;
Matches 144; Conservative 94; Mismatches 237; Indels 172; Gaps 24;
22 VSRAQQWVILSDENRNLQCEGCGYKVARLQKVETIQRVSEAYENLVKSSGKRALE 81
236 VKAAKDAADKMDIAEAKKREBAKTRFNTVRAMVVKAELELVKEE-----ANESNEEKI 290
82 KMNKLE---GIRSMHDPNDELERLETANKOLAEKEYEGSEDTRKTSIQLPKKNES 138
291 KQKEVSKKQAEATRLKIKTKDRKKAEEPEPEQAETK-KGSEAKQKAPELTKLEE 349
139 QREKEKLEELATARNSTNEDQRRHIEIRDOALSNAQAVKVLKEELKKQVYVD-----K 193
350 KRKAESSEKAAEAKQVDAEYVALEAK---IAELEVEQVLEKELXE-----IDERAKAK 402
194 VEMQOALVQQAACKEB-----OLEHRLRTRLERELESRIQ-----RQGNCP 240
403 LEEAEKATFAEKQVDAEVAPOAKIAELESQVH-RLQELKAGIDESOSDYKEGLRAP 451
241 TNVSEYNAALMELLSEKEERILALEADTKWE-----QKYLENVMRHFPALDAAATVA 294
462 LOSKLDTKKAKLSLEELSDKIDELDAEIAKLEVLQSESEDYAKEGP----- 508
295 AQDRTTVISHSPNTSVDTALBARIQKEEELMANKRCLDMEGIKTLHAQIIEKOAMIK 354
509 -----RAPLOKLDKAKKLSKLE-----LSDKIDELDAEIAKLEQLK 548
355 VLQORSRKPSQTEQLSCMRPAKSLMSISNAGSGLLSHSTLQSPIMEKRDDKSWKGS 414
549 DAEGNNVAYFKEGL-----EKTAAEKAELEKAZADLK----- 583
415 LGILLGGVRAEYVSTSPVPPSTPLLSAHSKGTGRDCSTQTERGTESNKTRAAPISV 474
584 -----KXVDEPTEPAPAK-----AENNNVED-----YKEGLEK----- 616
475 PAPVAAATAAATATATITTTTVAAPVAAAPVAAAAPAAAAPSPATAATAAASPAA 534
617 -----TIAAKKAELEKTEADLK-----AVNRPKPAFAPEPAP-----APEKPAKPA 661
535 AGQIPAAASVAGAAVAPSAASAAVAVAPAPVAPALVPVAPAPAAQASAPAQCA 594
662 APEKPA-----APEKPAPEKPAATRAP-EAPAEQKPAPOAPAPAPAPPEK- 710
595 PPSAPAVAPTAPT-----PTPAVAQ-AEVPASPATG 625
711 ----PABQPKPKTKDQQAEBDYARRPKPAPEKPAPEKPAPTPEKTKG 753

US-08-714-741-32

Query Match 7.9%; Score 263; DB 4; Length 8991;
Best Local Similarity 19.3%; Pred. No. 3.4e-08;
Matches 164; Conservative 83; Mismatches 298; Indels 306; Gaps 21;
2 PRAPFSASQYPVAPDFPAIVSRAQOMVEILSDENRN-----KPELKEIDSDSDYKGEFRAFLQSELD 52
S303 PAPKEKPAKPAPEP-----KPELKEIDSDSDYKGEFRAFLQSELD 52
53 LQKVETIQRVSEAYENLVKSSGKRALEKAMRNKLEHIEIRMDHFNDRDLERLETANKQ 112
S353 LSKLELSKIDELDAEIAKLEQLKAAE-----ENNVVEDYFKEGLEKTLIAAKKA 5403
113 LABKEGSDTKTISQISQFKNKESOREKELEAELATARNSTNEDQRRHIEIRDOALS 172
S404 ELEKT-----EADLKAVNBEKPAPEPSQPEKPAEAPAPAEQPTPEQPEKPAEQPAPA 5460
173 AQAVVLEELKXQVYVDYKVKQOALVQQAACEKQLEHRLRTRLERELESRIQ 232
S461 PQ-----PEKPAETPAPKPEKPAE---QPKAKPADQQAEBDYARRSSSEYVRLTQQ 5510
233 QROGNCOPT-----NVSEYNAALMELLSEKEERILALEADTKWEQKYLEENVMRH 284
S511 QPPKAEKPAKPKTKGSGALDQEAAPPHQVADLEKQITGPEIFLG----- 5556
285 PALDAATVAAQORDTTVISHSPNTSYDTALBARIQKEEELMANKRCLDMEGIKTLHA 344
S557 -GADPEADIAARENELAQAQAEKLAQKPTGLEKLLDS-----LDPGGKTQ----- 5599
345 QIIEKQAMIKVLOORSRKPSKTEQLSCMRPAKSLMSISN-----AGSGLLSHSTLQGS 399
S600 DELDKEAGEBAELDKKADLPNKVADLE-----KEISNLEILGGADSEDDTAALPNK 5651
400 PIMEEKRDDKSNK-CGLGILLGGYRAYVPS----- 430
S652 LAXKAELEKTKQELDAANEELGPDGDEETTPAPAPQEPQAPAPKPEQAPAPKPEOPA 5711
431 -----TPSEVP-----PSTPL----- 441
S712 PAPKPEOPAPAPKPEQAPAPAPKPEQAPKPEABEPTQPEKPAETKTRVRLKVAEFGVQ 5771
442 -----LSHSTKSGRDCSTQTERG-----TESNKTAAVAPISVPAPVAAA 481
S772 LRDAGSNVNGAYFKEGLEETTAEXEAGLGKAEADLKAVDEPETPAPAPAPAPAPAP 5831
482 ATAAATATAATITTTTVAAPVAAVAAAAPSPATAATAAATAAASVPAAGQIPAA 541
S832 APKPA-----PAPKPAAPAPAPAPKPAKPAKPAKPAKPAKPAKPAKPAKPA 5878
542 A----- 542

SULT 2

US-08-714-741-32

Sequence 32, Application US/08714741

Patent No. 6500613

GENERAL INFORMATION:

APPLICANT: Briles, David E.

APPLICANT: McDaniel, Larry S.

APPLICANT: Swiatlo, Edwin

APPLICANT: Yocher, Janet

APPLICANT: Crain, Marilyn J.

APPLICANT: Hollingshead, Susan

APPLICANT: Tart, Rebecca

APPLICANT: Brooks-Walter, Alexis

TITLE OF INVENTION: PNEUMOCOCCAL GENES, PORTIONS THEREOF,

TITLE OF INVENTION: EXPRESSION PRODUCTS THEREFROM, AND USES OF SUCH GENES,

TITLE OF INVENTION: PORTIONS AND PRODUCTS

NUMBER OF SEQUENCES: 47

CORRESPONDENCE ADDRESS:

ADDRESSEE: Curtis, Morris & Safford, P.C.

STREET: 530 Fifth Avenue

CITY: New York

STATE: New York

COUNTRY: U.S.

ZIP: 10036

COMPUTER READABLE FORM:

5879 APKPTPKTKLIDESDSEYAKGLRAPLQSLDTKAKLLKLELSGKIELDAEIXE 5938
543 -----SVASARAVASAAAAAV 560
5939 LEVOLDKDAEGNNVVEAYFKEGLEKTTAEKKAELKABDLKKAVDSEPTPAPAPAPAP 5998
561 QVAPAPAPAPAPALVPVAPAPAAQASAPAQTOAPTSAPAVAPTPTPTTAPVAAQAVPA 620
5999 APTPEAP 6058
621 -SPATGCPHR 630
6059 PAPAPAPKPKX 6069

SULT 3
-09-579-181-2
Sequence 2, Application US/09579181
Patent No. 6365372
GENERAL INFORMATION:
APPLICANT: Chivia, John
APPLICANT: Yaciuk, Peter
TITLE OF INVENTION: SNF2 Related CBP Activator Protein (SRCAP)
FILE REFERENCE: 16153-4247
CURRENT APPLICATION NUMBER: US/09/579,181
CURRENT FILING DATE: 2000-05-25
PRIOR APPLICATION NUMBER: 60/136,620
PRIOR FILING DATE: 1999-05-27
NUMBER OF SEQ ID NOS: 17
SOFTWARE: Patentin Ver. 2.0
SEQ ID NO 2
LENGTH: 2972
TYPE: PRT
ORGANISM: Human
-09-579-181-2

Query Match 7.5%; Score 252; DB 4; Length 2972;
Best Local Similarity 30.5%; Pred. No. 4.6e-08;
Matches 98; Conservative 34; Mismatches 103; Indels 86; Gaps 15;
430 STSPVPVPSTPLLSAHSKTSRDCSTQTERGTSNKTA-AVAPISVP-----475
1153 SSPWPIPNSSPLASPVSVSTVSLSSLSPTVPTLPAPASAPLTTPISAPLTVSASGPA 1212
476 -----APVAAA-----TAAAIT---ATAATITTTWAAAAPVAVA---507
1213 LLTSVTPLAPVWFAAPGPPSLQPSGASPSASALTGLATAPASAPLTTPISAPLTVSASGPA 1272
508 -----AAAAAPAAAAPSATAATAAAVSPAAAGQIPAAASVASAAAAPVAAAAAAV 560
1273 SHVPLNSTVAPACSPVLVPASALA-----SPFSPAPNPAQA-ASLLAPASSASQALAT 1326
561 QVAP-AAP-----APVPAPALVPVP-----APAAAQASAPAQTOAPTSAP-----AV 601
1327 PLAPMAAPQTAILAPSPAPPLAPLPLVAPSGAAPVLASSQTPVFNMAPSSTFGTSLASA 1386
602 APTPAPTPT--TPAVAQAEVPSAPATGCPGPHRLSIPSLTGNP--DKTDGPVFNHNTLERKT 657
1387 SPVPAPTPLVAFSSSTQTMFLPA-PVPSPLPSPASTQTLLALAPALAPTLLGSSSPQTLSLGT 1445
658 -----PIQLGQEPDAEMV 671
1446 GNPQGPPTQTLSLTPASSLV 1466

RESULT 4
3-09-579-181-1
Sequence 1, Application US/09579181
Patent No. 6365372
GENERAL INFORMATION:
APPLICANT: Chivia, John
APPLICANT: Yaciuk, Peter

TITLE OF INVENTION: SNF2 Related CBP Activator Protein (SRCAP)
FILE REFERENCE: 16153-4247
CURRENT APPLICATION NUMBER: US/09/579,181
CURRENT FILING DATE: 2000-05-25
PRIOR APPLICATION NUMBER: 60/136,620
PRIOR FILING DATE: 1999-05-27
NUMBER OF SEQ ID NOS: 17
SOFTWARE: Patentin Ver. 2.0
SEQ ID NO 1
LENGTH: 3118
TYPE: PRT
ORGANISM: Human
US-09-579-181-1

Query Match 7.5%; Score 252; DB 4; Length 3118;
Best Local Similarity 30.5%; Pred. No. 4.9e-08;
Matches 98; Conservative 34; Mismatches 103; Indels 86; Gaps 15;
430 STSPVPVPSTPLLSAHSKTSRDCSTQTERGTSNKTA-AVAPISVP-----475
1299 SSPWPIPNSSPLASPVSVSTVSLSSLSPTVPTLPAPASAPLTTPISAPLTVSASGPA 1358
476 -----APVAAA-----TAAAIT---ATAATITTTWAAAAPVAVA---507
1359 LLTSVTPLAPVWFAAPGPPSLQPSGASPSASALTGLATAPASAPLTTPISAPLTVSASGPA 1418
508 -----AAAAAPAAAAPSATAATAAAVSPAAAGQIPAAASVASAAAAPVAAAAAAV 560
1419 SHVPLNSTVAPACSPVLVPASALA-----SPFSPAPNPAQA-ASLLAPASSASQALAT 1472
561 QVAP-AAP-----APVPAPALVPVP-----APAAAQASAPAQTOAPTSAP-----AV 601
1473 PLAPMAAPQTAILAPSPAPPLAPLPLVAPSGAAPVLASSQTPVFNMAPSSTFGTSLASA 1532
602 APTPAPTPT--TPAVAQAEVPSAPATGCPGPHRLSIPSLTGNP--DKTDGPVFNHNTLERKT 657
1533 SPVPAPTPLVAFSSSTQTMFLPA-PVPSPLPSPASTQTLLALAPALAPTLLGSSSPQTLSLGT 1591
658 -----PIQLGQEPDAEMV 671
1592 GNPQGPPTQTLSLTPASSLV 1612

RESULT 5
US-09-252-991A-22853
Sequence 22853, Application US/09252991A
Patent No. 6551795
GENERAL INFORMATION:
APPLICANT: Marc J. Rubenfield et al.
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
FILE REFERENCE: 107196.136
CURRENT APPLICATION NUMBER: US/09/252,991A
CURRENT FILING DATE: 1999-02-18
PRIOR APPLICATION NUMBER: US 60/074,788
PRIOR FILING DATE: 1998-02-18
PRIOR APPLICATION NUMBER: US 60/094,190
PRIOR FILING DATE: 1998-07-27
NUMBER OF SEQ ID NOS: 33142
SEQ ID NO 22853
LENGTH: 399
TYPE: PRT
ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-22853

Query Match 7.5%; Score 250.5; DB 4; Length 399;
Best Local Similarity 25.4%; Pred. No. 4.6e-09;
Matches 107; Conservative 42; Mismatches 170; Indels 103; Gaps 13;
212 EQLHELETRLESLERLQIQGQNCQPTNVSYNAALMELLREKEERILALEADMTK 271
63 QQLSHSL-----VEHLGACK-----QALVDSEKULAKLEKQK 97


```

399  ---NQALRREERQA-----EREQEVIRHLEEB-----ORQLEI-----431
307 NTSYDTALEBARIKEREHEILMANKRCLDMEGRLKTLHAQILLEXDAMIKVLOQRSKBPSK 366
432  ---LOQLOEQALLLEYKRLQLEORQSERQRLQOQHAYLKSLQOQOQOQLQ 484
367 TEQLSCMRPAKSLMSISNAGSLHSSHTLTGSPIMBEKRDGSKWSGSLGILLOGDYRAE 426
485 KQOQOQLLPG-DRKPLVHYGRGMNPAKPAWAZEVEBTRMKNQONSLAKSKPGSTGPE 543
427 YVSTP--SPVPSSTLLSHSTGSDSCSTQTERGTESNKTAAVAPISVPAPVAAAATA 484
544  --PPIQASPGPG-PL-----SOTPMQRVPEQEGHKLVAHRVPLK---PYAAVPVR 593
485 AAITATAATITTTWAAAPVAVAAAAAPAAAAAPSPATAATAATAAVSPAAAGQIPAAASV 544
594  --SQSLQDQPTRNLAAPP-----ASHDPDPAIPAPTA---TPSARGAVIRQNSD 637
545 ASAAAVAPSAAAAAAVQVAPAAAPVP--APALVPVPAPAAAAQASAPAOQTAPSAPAVA 602
638 PTBEGPGFSPNPAWVRPDEAPKPVQRTSSITATLNTSGAGSRPAAQ-----AVR 689
603 PTAP-----TPTPAVAQAVPASPATGPGPHRLSISPLTCNPD 641
690 ARPSNSAWOILYORBERTKPK-----PGPPAOPPGP-----PNASNPND 731

```

```

RESULT 9
US-09-291-417D-15
; Sequence 15, Application US/09291417D
; Patent No. 6680170
; GENERAL INFORMATION:
; APPLICANT: FLOWMAN, GREGORY
; APPLICANT: MARTINEZ, RICARDO
; APPLICANT: WHYTE, DAVID
; TITLE OF INVENTION: STE20-RELATED PROTEIN KINASES
; FILE REFERENCE: 038602/0329
; CURRENT APPLICATION NUMBER: US/09/291,417D
; CURRENT FILING DATE: 1999-04-13
; PRIOR APPLICATION NUMBER: 60/081,784
; PRIOR FILING DATE: 1998-04-14
; NUMBER OF SEQ ID NOS: 155
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 15
; LENGTH: 1326
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-291-417D-15

```

Query Match	7.0%;	Score 234.5;	DB 4;	Length 1326;
Best Local Similarity	21.7%;	Fred. No. 2.4e-07;		
Matches 129;	Conservative 99;	Mismatches 237;	Indels 129;	Gaps 24
QY	68	ENLVKSSKR-EALEKAWRNKLEGEIRMHDFNRDLRRLRELTANKQLAKREYEGSESDTRK	126	
DB	247	EQLLQFPFIRDOPTQERVRIQLKDHID-----NSKKRGEKEETEYVSGSE-----	294	
QY	127	TISQLFPAKNKESQREHEKLEBAELATARNSTNEQDRRHLEIRQALSNQAQVKVLEELKK	186	
DB	295	-----EDDSHGEEGEPSSINNVPGESTLRREFLRQENKNSER-LKQQQLQQ	343	
QY	187	KQVYVDKVKMQQALVOLQAACEKREQLEHLRTLRLELESLRIQQQGNCGQPTNVSEY	246	
DB	344	QQ-QQDPEAHIKHLLHQRRRIEQKEERRRVEEQQRREQRKLQEK-----QQRRLED-	398	
QY	247	NAALMELLREKEERILALEADMTKWEQKYLEENVMRHFDALDAATVAQRDTTVIHSHP	306	
DB	399	-----MQLRREBERQA-----EREQYIHRRLSES-----QRQLEI-----	431	
QY	307	NTSYDTALSAIRIQEESBEILMANRCLDMEGRIKTYLHAQIITEKOMIKVLQQRSKESRK	366	
DB	432	-----LQCCLOSQALLVEYRKRLSEQRSESLRQLQGEHAYLSKLOQQQQQQQLQ	484	

367 TEQLSCMRPAKSLMSISNAGSLSSHTLGTSPIMEEKEDKSWKSGILGGLGGDYRAE 426
485 KQOQQQLPG-DRKPLHYGRGNPKADKPAWAEVBERITMNNQNSPLAKPGSGTGF 543
427 YVPSTP--SPVPSPTPLLSAHSKGTGSDCTQTERGTESNKTAAVAPISVPAPVAAAATA 484
544 --PPIQASGPPG-PL---SQTTPMQRVPBPQEGPHKSLVAHRVPLK---PYAAPVER 593
485 AAITATAATITTTWAAAPVAVAAAPAAAAPAAAAPSPATAATAAAVSPAAAGOIIPAAASV 544
594 ---SQSQDOQOTRNLAAFP-----ASHDDDDPAIPAPTA---TPSARGAVIRQNSD 637
545 ASAAVAPSAASAAAQVAPAAAPVP--APALVPVPAPAAQAQASAPAOQTOAPTSAPAVA 602
638 PTSEGGPGSPNPAWPNEDNEAPKVPQRTSSIIATALTNTSGAGGSRPAQ-----AVR 689
603 PTPAP-----TPTPAVAAQAEVPASPAATGPGPHLSIPSLTCHPD 641
690 APRSRNSAWQIYLORRAERGTGPD-----PGPPAQPQGP-----ENASSNPD 731

SULT 10
-09-252-991A-32957
Sequence 32957, Application US/09252991A
Patent No. 6551795
GENERAL INFORMATION:
APPLICANT: Marc J. Rubenfield et al.
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
FILE REFERENCE: 107196 136
CURRENT APPLICATION NUMBER: US/09/252,991A
CURRENT FILING DATE: 1999-02-18
PRIOR APPLICATION NUMBER: US 60/074,788
PRIOR FILING DATE: 1998-02-18
PRIOR APPLICATION NUMBER: US 60/094,190
PRIOR FILING DATE: 1998-07-27
NUMBER OF SEQ ID NOS: 33142
SEQ ID NO 32957
LENGTH: 316
TYPE: PRT
ORGANISM: Pseudomonas aeruginosa
-09-252-991A-32957

Query Match 6.9%; Score 231; DB 4; Length 316;
Best Local Similarity 29.2%; Pred. No. 6.7e-08;
Matches 107; Conservative 29; Mismatches 129; Indels 102; Gaps 15;
302 ISHSPNTSYDTALEARIQKBEIEILMANKRCLDMEGRIKTLHAQI-----TEKDA 351
6 VSRBNKESRMAGKKSEKSSWI-----GEIEKYSRQIMLAGLAYSKVSKDG 54
352 MIKVLQORSRKEPKTEQLSCMRPAKSLMSISNAGSLSSHTLGTSPIMEEKRD--- 407
55 --SKLFETLVKDGKAE-----KEAKSDVDAQVGAAKASARSASKV-DEVRDALG 103
408 -----DKSWKSGILGGLGGDYRAEYVPSTPSPVPPPTPLLSAHSKGTGSDCTQTE 458
104 KWSLEEAFFDXRLNSAISL-----GVPSRNEVKELHSKVDI--LTKQIE 146
459 RGT-----ESNKTAAVAIISVPAPVAAAATAAATAAATATATITTT 497
147 KLTGVSVKPAKAAKPAKPAKPAKPAKPAKPAKPAKPAKPAKPAKPAKPAKPAKPA 203
498 MWAAAPVAAAAPAAAAAPSPATAATAAASVPAAGOTPAAASVAAAAPAAAAA 557
204 KTAAPKPAKPAKPAKPAKPAKPAKPAKPAKPAKPAKPAKPAKPAKPAKPAKPA 260
558 AAQVAPAPAPVAPALVPVAPAPAAQASAPAOQTOAPTSAPAVPTPTTPTTAAQAE 617
261 A-----KPAKKPA-----AKKPAKPAKPAKPAKPAKPAKPAKPAKPAKPAK 307
618 VPASPAT 624
|||:::

Db 308 APATPSS 314
RESULT 11
US-08-328-254-6
; Sequence 6, Application US/08328254
; Patent No. 5710022
; GENERAL INFORMATION:
; APPLICANT: Zhu, Xueliang
; APPLICANT: Lee, Wen-bwa
; TITLE OF INVENTION: A No. 5710022el Nuclear Mitotic Phosphoprotein
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Campbell and Flores
; STREET: 4370 La Jolla Village Drive, Suite 700
; CITY: San Diego
; STATE: California
; COUNTRY: USA
; ZIP: 92122
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/328,254
; FILING DATE: 24-OCT-1994
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/141,239
; FILING DATE: 22-OCT-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Campbell, Cathryn A.
; REGISTRATION NUMBER: 31,815
; REFERENCE/DOCKET NUMBER: P-CJ 1191
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 535-9001
; TELEFAX: (619) 535-8949
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2482 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-328-254-6

Query Match 6.8%; Score 226.5; DB 1; Length 2482;
Best Local Similarity 21.4%; Pred. No. 1.8e-06;
Matches 139; Conservative 100; Mismatches 233; Indels 179; Gaps 26;
QY 26 QQVVEILSDENRNLQBLE-----GCYEKVARLOKVETEIQVSEAVENLVKSSKRE 78
1948 QDTLEVLSQSYKLENELELTTKDKMSFVEKNKMTAKETELQ--EMHEMAQKTAELQE 2005
QY 79 ALEKAMNKLEGEIRMHED-----FNRLRERLETANKQLAEKEYEGSD- 123
2006 EL-SGEKNRLAGELQLLELTKSKQOLKELTLENSELKSLDCKMDQVKEKGKVRBEI 2064
QY 124 -----TRKTIISOLFANKEQREKLEAEATARSTNEDQRRHIEI---RDQAL 170
2065 AEYQLRLHEAKKHQALLDNTNKQYVEIQTYYREKLTSECECLSSQKLEIDLLKSKKEEL 2124
QY 171 SNAQKVVVLEELKKQV-----YVDKV-----EKMQQALVQLQAACEKBPOLSHRLTRL 222
2125 NNSLKATTTQLEELKTKMDNKLKYNQKXENRAQCKMLLKSKNQLEEKKEIQLKEL 2184
QY 223 ERELESRIQORQNCQOPTNVSEYNA--AALMELLREKEERILALEADMTKWQKYLEEN 280
2185 S-QLOAQEKQKTQVMDTKVDELTTTEIKELKTELEK-----TKBADEYLDKY 2232
QY 281 VWRHFALDAATAVAQDDTTTIVSHSPNTSYDTALEARIQKBEIEILMANKRCLDMEGRIK 340
2233 C-----SLLISHE-----KLEKAKEM-----LETQVA 2254

```

341 TLHAQIIEKDA-----MIKVLQQRKRPEPSKTEBOLSCMRPAKSLMSISNAGSGLLSHSSTL 396
      ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
2255 HLCQSQXKQDSRGSLPLAGVVPVPGSPITPSVTE-----KELSSGQWKAGSKRQRSSGI 2306
      ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
397 TGSPIWEEKRODKSWKGSGLGILLGGDYRAEYVVPSTSPVPPSTPLLSAHSKTSGRDCSTQ 456
      ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
2307 -----WENG-----GG-----PTPATPSPSKSKKAVMSGIHPAE 2337
      ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
457 TERGTESNKTAAVAPISVPAPVAAAATAAAITATAATI---TTTWVAAAPVAVAAAAAPAA 514
      ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
2338 DTGEGTE-----FPREGLPVVKVKGADIPGTGTSYILERTTM----- 2375
      ::::: ::::: ::::: ::::: ::::: ::::: :::::
515 AAAPSPATAATAAAVAPAAAGIIPAAASVASAAAAPSAAAAAAQAQVAPAPAPVAPA 574
      ::::: ::::: ::::: ::::: ::::: ::::: :::::
2376 AIRTSP-RLAQKALSLSLSKENLAES-----SKPTAGSRSQKVKVQSRFVDSGT 2428
      ::::: ::::: ::::: ::::: ::::: :::::
575 LVVPVAPAAAQAQAPAOQAPTQAPAVAPTPPATPTTAVAAQAEVAPSPATG 625
      ::::: ::::: ::::: ::::: ::::: :::::
2429 ILREP-----TTKSVFV-----NNLPRSPPTDSPREGLRVKGRGLVPSPKAG 2470
      ::::: ::::: ::::: ::::: ::::: :::::
SULT 12
09-963-137-184
Sequence 184, Application US/09963137
Patent No. 6596036
GENERAL INFORMATION:
APPLICANT: Pedersen, Finn S
APPLICANT: Sorensen, Annette B
APPLICANT: Hernandez, Javier Martin
APPLICANT: Nielsen, Anne A
APPLICANT: Moeving, Helle
TITLE OF INVENTION: NOVEL COMPOSITIONS AND METHODS FOR LYMPHOMA AND LEUKEMIA
FILE REFERENCE: A-70981/RMS/DCF
CURRENT APPLICATION NUMBER: US/09/963,137
CURRENT FILING DATE: 2001-09-24
PRIORITY APPLICATION NUMBER: US 09/668,644
PRIORITY FILING DATE: 2000-09-22
PRIORITY APPLICATION NUMBER: US 09/905,390
PRIORITY FILING DATE: 2001-07-13
PRIORITY APPLICATION NUMBER: US 09/905,491
PRIORITY FILING DATE: 2001-07-13
PRIORITY APPLICATION NUMBER: US 09/962,929
PRIORITY FILING DATE: 2001-09-24
PRIORITY APPLICATION NUMBER: US 09/962,854
PRIORITY FILING DATE: 2001-09-24
PRIORITY APPLICATION NUMBER: US 09/962,916
PRIORITY FILING DATE: 2001-09-24
PRIORITY APPLICATION NUMBER: US 09/962,855
PRIORITY FILING DATE: 2001-09-24
NUMBER OF SEQ ID NOS: 215
SOFTWARE: PatentIn version 3.1
SEQ ID NO 184
LENGTH: 756
TYPE: PRT
ORGANISM: Mus musculus
-09-963-137-184
Query Match 6.7%; Score 225.5; DB 4; Length 756;
East Local Similarity 30.1%; Pred No. 4.7e-07;
Matches 87; Conservative 37; Mismatches 108; Indels 57; Gaps 11;
374 RPAKSLMSIS-NAGSGLLSHSSTLTGSPIMEEKRDDKSKWKSGLGILLGGDYRAEYVVPST- 431
      ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
34 KEAPAAQAVSTGSDAGAPTDSAMLTDS--QSDAGEDTAPGTPTDQLQDPELFEAPAVR 91
      ::::: ::::: ::::: ::::: ::::: :::::
432 -----PSPVPPSTPLLSAHSKTSGRDCSTQTERGTESNKTAAVAPISVPVAAAAATAA 485
      ::::: ::::: ::::: ::::: ::::: :::::
92 ADPDGGAAPVAPATPAES--ESEGSD-----PAAPHASEAVPATTA 131
      ::::: ::::: ::::: ::::: :::::
486 AITATAATITTTWAAAPVAVAAAAAPAAAAAP-----SPATAAATAAAVSPAAAGIIPAA 541
      ::::: ::::: ::::: ::::: :::::
132 ESASGAAPVTOVEPAAAVS-ATLAPAPAAAPITPKPTTAVPSSAAHP-AAGAVPGA 189
      ::::: ::::: ::::: ::::: :::::

```

[illegible]

SULT 14
TUS95-16216-1
Sequence 1 Application PC/TUS9516216
GENERAL INFORMATION:
APPLICANT: Yes, Timothy J.
APPLICANT: Rattnar, Jerome B.
TITLE OF INVENTION: Nucleic Acid Encoding a Transiently
Expressed Kinetochores Protein, and Methods of Use
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:

Query Match	6.7%;	Score 225;	DB 5;	Length 3248;
Best Local Similarity	21.4%;	Pred. No. 3.2e-06;		
Matches 110:	Conservative 106;	Mismatches 167;	Indels 132;	Gaps 21;

Qy	20	AVIRSAQOMVEIISDENMLRQBELCEYKVARLOK-----VETRIQVRSEAYE	68
Db	2103	AVYKERTELLQTLSSDVSELLKDKXQHLQKOSLEKDSQALSTKCELENOIAGLNNKEKE	2162
Qy	69	NLVXSSSK-----REALERAMENKLEGBIR-----RMEDFNRR---DLRE	104
Db	2163	LLVXSEESIQARLSSESDYBKLVNSVKALBAALVEKGEFALRLSSQTQEEVHQLRRGIEKLVR	2222
Qy	105	RLFTANKQ---LAEK-----EYGESEDETRKTSIQLFAKNKESQR-----BKEKLEASLAT	151
Db	2223	RIEADKKQLHIAEKUKBERRENDLSUKQVEMLERELQWSENGELVILDAENSKAEVET	2282
Qy	152	ARSTNEPDRRHIEIRDQALSNAAQAVVKEELBKQVYVDVKVRM-----OQA	200
Db	2283	LKTOIEEMARSLKIFELDLVTLRSEKENLTQIOEQKQGLSELDKLLSSFKSLLEKEQA	2342
Qy	201	LVQLOQ-----AACERB-----QLEHRLRTRLE	225
Db	2343	EIOKKEESTFAVEMLOQLKELNBEAVALCGDQOIKMTAKTQBSLDPPIEERHQLRNSTEKL	2402
Qy	226	LESRLIOQRQNC---OPTNVSEYNAAL-----MELLRKEERILALEADMTKWE	273
Db	2403	RARLEAEKKQLCVLOLKESEHHADLKGRVENLERELEIARTNQEH--AALEAENSKGE	2461
Qy	274	QKYLEENV-----MRHFALDAAA-----TVAAQRDITTVISHSP---NTSYOTALEAR	317
Db	2462	VETLKAKTEGQTSQRLGLELDVVTIRSEKENLTNLOXEQERISELEINSSPENILO--	2519
Qy	318	IOKEBEETILMANKECLDMEGRIKYLTHAQIIIEKDAMIKVLQQRSR--KEPSKTEQULSC--MR	374
Db	2520	-EKEGBQVMKRSSTANE--MLQTLQELNURVAALHNDQEAACKAQNLSQVCECLE	2577
Qy	375	PAKSLMSISNAGSGLSHSSTLTGSPIMEKKDDK	409
Db	2578	KAOILOGDEAKNNYIVLOSXVKG--LIQEVBDGK	2610

RESULT 15
US-09-489-039A-7502
: Sequence 7502. Application US/09489039A

Query Match 6.7%; Score 223.5; DB 4; Length 1079;
Best Local Similarity 22.3%; Pred. No. 1e-06;
Matches 145; Conservative 65; Mismatches 266; Indels 175; Gaps 25;

Qy	1	MRAQPSSASVQVPAD-----PPAIVSRAQQWVTILSDENLNTQLQEGLCGCVKVARLQK	55
Dd	536	MDAPPAFMLSEPAAPVAAPAAAPAP--GLLSRPFPSALKNIISGAEE-----AKP	589
Qy	56	VETETQRVSAYENVLVSSKKRKALPKAMNKLEGISIRMHDFNDRLERLETANKOLAE	115
Dd	590	AEOVQKEKAEE-----KPERQORERKPRANNRRDRNRDNEDNRDNDNRD-----	636
Qy	116	KEYEGSDTRXTVISQLFAKNKSQREKEKLAELATARSTNEDQRRHIEIRDOALSNAQA	175
Dd	637	-----NNTP-----ADNAECGRDPRESRE-----ENRNREKPKSONVEARD-----	673

```

Query Match          6.7%; Score 225; DB 5; Length 3248;
Best Local Similarity 21.4%; Pred. No. 3.2e-06;
Matches 110: Conservative 106; Mismatches 167; Indels 13

```

```

176 KVKLEBEELKKQVYDVKEKQCALVOLQAACEKEQLEHLRTRLERELESRIQORQ 235
574 -----VRQTSDDAEK-----AKSRDEQOPRRRTA--RRSDDKROAQOE 711
236 GNCQPTN---VSEYNAAALMELLREKEERILALEADMTKEQKYLEENVMRHFALDAAAT 292
712 AKAQTRREEPVQOE-----TEQEEERVQTL-----PRKPRQLAQKVR----- 747
293 VAAQRDTTVISHSPNTSYDTALEARIQKEBEEILMANKRCLDMEGRIKTLHAQIIEKDAM 352
748 -----VESAVVEPVAEIIVPEAVV 765
353 IKVLQQRSRKEPSKTSQLSQMRPAKSLMSISNAGSGLLSHSTLTGSPIMEKRDOKSWK 412
766 AEVIAPES--EPVKAE-----LPA-GVESVADQDEN--GESREANGP-----RRSRR 808
413 GSLGILLGSD---YBAEYVPSTPSPVPSTPLLSAHSKTGS-----RDCSTQTRRG 460
809 SPRHLRVSGQRRRYRDERYP-TQSPMPUTVACASPEMASGKWIRYPVWRPQDOQPREEV 867
461 TESNKTAAVAPISVPAPVAAAAATAAATAATITTTMVAAPVAVAAAAAPAAAAAPSP 520
868 QVQDASVAKTVEAVAAPVAVETVTAAPVTVEPATMEPVTAEPVVV---EPVAAAAEPLF 923
521 ATAA--ATAAANSPAAAGQIPAAASVASAAVAPSAASAAAAVQVAPAPAPVPAPAL-VP 577
924 VDAAEVVAAPAAVEPAPPOEPVTEAPAVEAPQAIAPVTLDAEPVVVEPEAVETTPVVAAPVE 983
578 VPAPAAQAQASAPAQTSAPAVAPTPTPTPAVAQAEVPASPATGPGP 628
984 TIAPVAETVEQAPVTEAPAPAEPVKAEPVSKPVVWVAGERHATAPWTRAPAP 1034

```

arch completed: June 16, 2004, 19:13:23
time : 31.0562 secs

[illegible]

ALIGNMENTS

RESULT 1	US-10-204-887-87	Sequence 87, Application US/10204887
		Publication No. US20030124569A1
		GENERAL INFORMATION:
	APPLICANT:	INCYTE GENOMICS, INC.
	APPLICANT:	PANZER, Scott R.
	APPLICANT:	SPIRO, Peter A.
	APPLICANT:	BANVILLE, Steven C.
	APPLICANT:	SHAH, Purvi
	APPLICANT:	CHALUP, Michael S.
	APPLICANT:	CHANG, Simon C.
	APPLICANT:	CHEN, Alice
	APPLICANT:	D'SA, Steven A.
	APPLICANT:	AMEHEY, Stefan
	APPLICANT:	DAHL, Christopher R.
	APPLICANT:	DAM, Tam C.
	APPLICANT:	DANIELS, Susan E.
	APPLICANT:	DOUFOR, Gerard E.
	APPLICANT:	FLORES, Vincent
	APPLICANT:	FLORES, Willy T.
	APPLICANT:	GREENAWALT, Lila B.
	APPLICANT:	HILLMAN, Jennifer L.
	APPLICANT:	JONES, Anissa L.
	APPLICANT:	LIU, Tommy F.
	APPLICANT:	ROSEBERY, Ann M.
	APPLICANT:	ROSEN, Bruce H.
	APPLICANT:	RUSSO, Frank D.
	APPLICANT:	STOCKREHER, Theresa K.
	APPLICANT:	DAFFO, Abel
	APPLICANT:	WRIGHT, Rachel J.
	APPLICANT:	YAP, Pierre E.
	APPLICANT:	YU, Jimmy Y.
	APPLICANT:	BRADLEY, Diana L.
	APPLICANT:	BRATCHER, Shawn R.
	APPLICANT:	CHEN, Wensheng
	APPLICANT:	COHEN, Howard J.
	APPLICANT:	HODGSON, David M.

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

protein - protein search, using sw model

June 16, 2004, 19:10:47 ; Search time 65.1895 Seconds
(without alignments)
2917.113 Million cell updates/sec

File: US-09-332-063-2

fect score: 3347
 nence: 1 MPRAPSSASYQVPADPPA.....KTPIQILGQEPDAEWVEYLI 675

```
oring table: BLOSUM62
Gapop 10.0 , Gapext 0.5
```

arched: 1158786 seqs. 281726120 residues

total number of hits satisfying chosen parameters: 1158786

```
nimum DB seq length: 0
ximum DB seq length: 2000000000
```

```
st-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 su
```

Database : Published Applications AA: *

```

1: /csm2_6/ptodata/2/pubpaa/us07 PUBCUMB pep.*
2: /csm2_6/ptodata/2/pubpaa/PCT_NEW PUB pep.*
3: /csm2_6/ptodata/2/pubpaa/us06_NEW PUB pep.*
4: /csm2_6/ptodata/2/pubpaa/us06 PUBCUMB pep.*
5: /csm2_6/ptodata/2/pubpaa/us07 NEW PUB pep.*
6: /csm2_6/ptodata/2/pubpaa/PCTUS PUBCUMB pep.*
7: /csm2_6/ptodata/2/pubpaa/us08 NEW PUB pep.*
8: /csm2_6/ptodata/2/pubpaa/us08 PUBCUMB pep.*
9: /csm2_6/ptodata/2/pubpaa/us09A PUBCUMB pep.*
10: /csm2_6/ptodata/2/pubpaa/us09B PUBCUMB pep.*
11: /csm2_6/ptodata/2/pubpaa/us09C PUBCUMB pep.*
12: /csm2_6/ptodata/2/pubpaa/us09C_NEW PUB pep.*
13: /csm2_6/ptodata/2/pubpaa/us10A PUBCUMB pep.*
14: /csm2_6/ptodata/2/pubpaa/us10B PUBCUMB pep.*
15: /csm2_6/ptodata/2/pubpaa/us10C PUBCUMB pep.*
16: /csm2_6/ptodata/2/pubpaa/us10C_NEW PUB pep.*
17: /csm2_6/ptodata/2/pubpaa/us10C_NEW PUB pep.*
18: /csm2_6/ptodata/2/pubpaa/us60 PUBCUMB pep.*

```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

result No.	Score	Query			ID	Description
		Match	Length	DB		
1	1447	43.2	608	14	US-10-204-987-97	Sequence 87, Appl
2	1400	41.8	862	14	US-10-236-117-2	Sequence 2, Appl
3	1090.5	32.6	772	14	US-10-236-117-4	Sequence 4, Appl
4	1037	31.0	467	15	US-10-064-466-52	Sequence 52, Appl
5	280.5	8.4	1259	14	US-10-250-715-8	Sequence 8, Appl
6	278	8.3	208	9	US-09-864-711-36456	Sequence 36456, A
7	275	8.2	1701	14	US-10-171-311-83	Sequence 83, Appl
8	265.5	7.9	2846	14	US-10-184-644-169	Sequence 169, App
9	265.5	7.9	2846	14	US-10-184-634-169	Sequence 169, App
10	265.5	7.9	2846	14	US-10-063-685-37	Sequence 37, Appl
11	264.5	7.9	2773	14	US-10-184-644-149	Sequence 149, App
12	264.5	7.9	2773	14	US-10-184-634-149	Sequence 149, App
13	264.5	7.9	2773	14	US-10-063-685-37	Sequence 33, Appl
14	263.5	7.7	1965	15	US-10-369-493-3279	Sequence 3279, Ap
15	258	7.7	980	14	US-10-029-386-33686	Sequence 33686, A

APPLICANT: LINCOLN, Stephen E.
TITLE OF INVENTION: SECRETORY MOLECULES
FILE REFERENCE: PT-1134 PCT
CURRENT APPLICATION NUMBER: US/10/204,887
CURRENT FILING DATE: 2002-08-21
PRIOR APPLICATION NUMBER: 60/185,215; 60/185,216; 60/205,232; 60/205,287;
60/205,324; 60/205,286
PRIOR FILING DATE: 2000-02-24; 2000-02-24; 2000-05-16; 2000-05-17; 2000-05-17;
2000-05-17; 2000-05-17
NUMBER OF SEQ ID NOS: 159
SOFTWARE: PERL Program
SEQ ID NO 87
LENGTH: 608
TYPE: PRT
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc feature
LOCATION: 18, 388
OTHER INFORMATION: unknown or other
-10-204-887-87
Query Match 43.2%; Score 1447; DB 14; Length 608;
Best Local Similarity 62.1%; Pred. No. 2.4e-74;
Matches 293; Conservative 79; Mismatches 84; Indels 16; Gaps 6;
2 PRAPSSASQVPADPFAIVSRAQQWVILSDENRNLRQELGGCEKVARLQKVETEIQ 61
70 POPPPASQOQLGPDFAIVSRAQQWVILSDENRNLRQELGGCEKVARLQKVETEIQ 129
62 RVSEAYENLVKSSSKRALEKAMRNKLEGBIRRMHDFNDRLESLRLESLRLESLR 121
130 RISEAYESLVKSTTKRESLDKAMRNKLEGBIRRMHDFNDRLESLRLESLRLESLR 189
122 EDTRKTIQSOLFAPNKESQREKELEAEALATARNSTNEDORRHIEIRDOALSNAQAKVVKLE 181
190 ED-KAAGHVASQNKFEKLEKLEAEALATARNSTNEDORRHIEIRDOALSNAQAKVVKLE 248
182 EELKKQVYVDKVEKQQAALVOLQAACEKEQLEAEALATARNSTNEDORRHIEIRDOALSNAQAKVVKLE 241
249 EELREKQAYVEKVKLQQAALVOLQAACEKEQLEAEALATARNSTNEDORRHIEIRDOALSNAQAKVVKLE 308
242 NVSEYNAALMELREKBERILAEADMTKWEQKYLEENVMRHFDALDAATVAARDTIV 301
309 NMPEYNAPALLEVREKBERILAEADMTKWEQKYLEENVMRHFDALDAATVAARDTIV 368
302 ISHSPNTSY-DTAEARIQKEEBEILMANKRCLDMEGRIKTLHAQIIIEKDAMIKVLQORS 360
359 INHSPNGSYGESSLEAHINWEEBEVQANRRCCQDMETYNKHLAKIIEKDAMIKVLQORS 428
361 RXEBSKTEQLSCMRPAKSLMSISNAGSGLLSHSSTLTGSPIMEEKDDKSWKSGILILG 420
429 RKDAGKTDS--SSLRPARSVPSIA-AATGTHSRQTSLSLTSSQLAEKKEBKTKWKSIGLILG 486
421 GDYRAEYVPSPTSPVPSPTPL-----SAHSKTSGRDCSTOTTERGT 462
487 KEHHEH--ASAPLLPPTSLSSIASTTAASSNAHAKTSGSKDSTQTDKSAE 536

APPLICANT: Yoshiaki Takai
TITLE OF INVENTION: Exocrine gland tight junction-constituting protein JEAP family
FILE REFERENCE:
CURRENT APPLICATION NUMBER: US/10/298,417
CURRENT FILING DATE: 2003-01-22
PRIOR APPLICATION NUMBER: JP 2001-352241
PRIOR FILING DATE: 2001-11-16
NUMBER OF SEQ ID NOS: 10
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 2
LENGTH: 882
TYPE: PRT
ORGANISM: Mus musculus
US-10-298-417-2
Query Match 41.8%; Score 1400; DB 14; Length 882;
Best Local Similarity 58.2%; Pred. No. 1.8e-71;
Matches 298; Conservative 80; Mismatches 103; Indels 32; Gaps 11;
QY 1 MPRAQP--SSASYQVPADP-----FAIVSRAQQWVILSDENRNLRQELGGCEKVA 51
DB 333 LPLPLPTSLAASQPLPASPNQQLGPDFAIVSRAQQWVILSDENRNLRQELGGCEKVA 392
QY 52 RLQKVETEIORVSEAYENLVKSSSKRALEKAMRNKLEGBIRRMHDFNDRLESLRLESLR 111
DB 393 KHKFEKELQSISEAYESLVKSTTKRESLDKAMRNKLEGBIRRMHDFNDRLESLRLESLR 452
QY 112 QLAKEYESGEDTRKTIQSOLFAPNKESQREKELEAEALATARNSTNEDORRHIEIRDOALS 171
DB 453 QLSREYDGHED-KAAESHYVSQNKFEKLEKLEAEALATARNSTNEDORRHIEIRDOALS 511
QY 172 NAGAKVVKLEKQVYVDKVEKQQAALVOLQAACEKEQLEAEALATARNSTNEDORRHIEIRDOALS 231
DB 512 NAGARVFKLEKQVYVDKVEKQQAALVOLQAACEKEQLEAEALATARNSTNEDORRHIEIRDOALS 571
QY 232 QORQGNQCPNTSEYNAALMELREKBERILAEADMTKWEQKYLEENVMRHFDALDAATVA 291
DB 572 QOKHGTGPPVSLPECNAPALMELVREKBERILAEADMTKWEQKYLEENVMRHFDALDAATVA 631
QY 292 TVAQRDTTVISHSPNTSY-DTAEARIQKEEBEILMANKRCLDMEGRIKTLHAQIIIEKD 350
DB 632 AATAERTTISNHRNGSYGESSLEAHINWEEBEVQANRRCCQDMETYNKHLAKIIEKD 691
QY 351 AMIKVLQORSKPEKTEQLSCMRPAKSLMSISNAGSGLLSHSSTLTGSPIMEEKDDK- 409
DB 692 AMIKVLQORSKPDAGKTDSSAS-LRPARSVPSIA-AATGTHSRQTSLSLTSSQLAEKKEBK 749
QY 410 SWKSGILILG-----GDYRAEYVPSPTSPVPSPTPL-----PVPSPPTPLLSA-HSKTSGRDCSTOTTERGT 461
DB 750 TWKSGIGFLGKQHQQAASAPLPTTPASALSUPASTTSASSTHAKTSGKDSSTQTDKST 809
QY 462 E-----SNKTAAPVAPISVPAPVAAAAATAA 485
DB 810 ELFWPSNASLPSRGRSLSTAFSPNSPILKHPAAKGA 843
RESULT 3
US-10-298-417-4
Sequence 4, Application US/10298417
Publication No. US20030124603A1
GENERAL INFORMATION:
APPLICANT: Miyuki Nishimura
APPLICANT: Mayumi Asano
APPLICANT: Yuichi Ono
APPLICANT: Koji Morimoto
APPLICANT: Masakazu Takeuchi
APPLICANT: Yoko Inoue
APPLICANT: Toshio Imai
TITLE OF INVENTION: Exocrine gland tight junction-constituting protein JEAP family
FILE REFERENCE:
CURRENT APPLICATION NUMBER: US/10/298,417
CURRENT FILING DATE: 2003-01-22

2709 ACTACAAAAA
553 SAAAAAVQVAPAPAPVAPALVPVAPAPAAQASAPACTOATTSAPAVAPTPTPTPA 612
2769 AAAAAA
613 VQAQEVPA 623
2829 AAAAAA 2839

SULT 10
-10-063-685-37
Sequence 37, Application US/10063685
Publication No. US20030180909A1
GENERAL INFORMATION:
APPLICANT: Eaton Dan L.
APPLICANT: Filvaroff, Ellen
APPLICANT: Gerritsen, Mary E.
APPLICANT: Goddard, Audrey
APPLICANT: Godowski, Paul J.
APPLICANT: Grimaldi, Christopher J.
APPLICANT: Gurney, Austin L.
APPLICANT: Watanabe, Colin K.
APPLICANT: Wood, William I.
TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
FILE REFERENCE: P3230R1C1
CURRENT APPLICATION NUMBER: US/10/063.685
CURRENT FILING DATE: 2002-05-08
Prior Application removed - See Palm or File Wrapper
NUMBER OF SEQ ID NOS: 170
SEQ ID NO 37
LENGTH: 2846
TYPE: DNA
ORGANISM: Homo Sapien
-10-063-685-37

Query Match 7.9%; Score 265.5; DB 14; Length 2846;
Best Local Similarity 42.4%; Pred. No. 2.7e-06;
Matches 81; Conservative 11; Mismatches 84; Indels 15; Gaps 2;
448 TGRDCSTQT-----ERTESNKTAATAVAPISVPAPVAAATAA-----AITATTA 492
2649 TGGTCCTTTCTTCCCATCTCTGTACACATTTTAAATAAATAAGGTTGGCTTCTGA 2708
493 TTTTWWAAAVVAVAAAPAAAAAAPSATAATAAVSPAAAGQIPAAASVASAAP 552
2709 ACTACAAAAA 2768
553 SAAAAAVQVAPAPAPVAPALVPVAPAPAAQASAPACTOATTSAPAVAPTPTPTPA 612
2769 AAAAAA 2828
613 VQAQEVPA 623
2829 AAAAAA 2839

RESULT 11
-10-184-644-149
Sequence 149, Application US/10184644
Publication No. US2003004930A1
GENERAL INFORMATION:
APPLICANT: Baker, Kevin P.
APPLICANT: Chen, Jian
APPLICANT: Desnoyers, Luc
APPLICANT: Goddard, Audrey
APPLICANT: Godowski, Paul J.
APPLICANT: Gurney, Austin L.
APPLICANT: Pan, James
APPLICANT: Smith, Victoria
APPLICANT: Watanabe, Colin K.

APPLICANT: Wood, William I.
APPLICANT: Zhang, Zemin
TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
FILE REFERENCE: P3430R1C27
CURRENT APPLICATION NUMBER: US/10/184.644
CURRENT FILING DATE: 2002-06-28
Prior Application removed - See File Wrapper or Palm
NUMBER OF SEQ ID NOS: 612
SEQ ID NO 149
LENGTH: 2773
TYPE: DNA
ORGANISM: Homo Sapien
US-10-184-644-149

Query Match 7.9%; Score 264.5; DB 14; Length 2773;
Best Local Similarity 43.4%; Pred. No. 3e-06;
Matches 79; Conservative 11; Mismatches 85; Indels 7; Gaps 2;
QY 449 GSRDCSTQTERGTESNKTAATAVAPISVPAPVAAATAAATAATAATTTTWWAA-----AP 503
DB 2594 GTTTCATTTTGTCA--TGCAATGTAGGATTCGCAATTAATCTTTAGAGGATGAA 2651
QY 504 VAVAAAAAPAAAAAPSATAATAAVSPAAAGQIPAAASVASAAPAAAAVQVAA 563
DB 2652 AATAAAAAA 2711
QY 564 PAAAPVPAPALVPVAPAPAAQASAPACTOATTSAPAVAPTPTPTPAVQAQEVPA 623
DB 2712 AAAAAA 2771
QY 624 TG 625
DB 2772 AG 2773

RESULT 12
US-10-184-634-149
Sequence 149, Application US/10184634
Publication No. US2003006868A1
GENERAL INFORMATION:
APPLICANT: Baker, Kevin P.
APPLICANT: Chen, Jian
APPLICANT: Desnoyers, Luc
APPLICANT: Goddard, Audrey
APPLICANT: Godowski, Paul J.
APPLICANT: Gurney, Austin L.
APPLICANT: Pan, James
APPLICANT: Smith, Victoria
APPLICANT: Watanabe, Colin K.
APPLICANT: Wood, William I.
APPLICANT: Zhang, Zemin
TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
FILE REFERENCE: P3430R1C217
CURRENT APPLICATION NUMBER: US/10/184.634
CURRENT FILING DATE: 2002-06-28
Prior Application removed - See File Wrapper or Palm
NUMBER OF SEQ ID NOS: 612
SEQ ID NO 149
LENGTH: 2773
TYPE: DNA
ORGANISM: Homo Sapien
US-10-184-634-149

Query Match 7.9%; Score 264.5; DB 14; Length 2773;
Best Local Similarity 43.4%; Pred. No. 3e-06;
Matches 79; Conservative 11; Mismatches 85; Indels 7; Gaps 2;
QY 449 GSRDCSTQTERGTESNKTAATAVAPISVPAPVAAATAAATAATAATTTTWWAA-----AP 503
DB 2594 GTTTCATTTTGTCA--TGCAATGTAGGATTCGCAATTAATCTTTAGAGGATGAA 2651

504 VAVAAAAAPAAAAAPSPATATAATAAVSPAAAGQIPAAASVASAAVAPSAASAAVQVA 563
2652 AATAAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAA 2711
564 PAAPAPVPALVVPAPAAQASAPAOQAPTSAPAVAPTPTPTPAVAQAEVPSA 623
2712 AAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAA 2771
624 TG 625
2772 AG 2773

SULT 13
-10-063-685-33
Sequence 33, Application US/10063685
Publication No. US20030180909A1
GENERAL INFORMATION:
APPLICANT: Eaton, Dan L.
APPLICANT: Filvaroff, Ellen
APPLICANT: Gerritsen, Mary E.
APPLICANT: Goddard, Audrey
APPLICANT: Goddard, Audrey
APPLICANT: Godowski, Paul J.
APPLICANT: Grimaldi, Christopher J.
APPLICANT: Gurney, Austin L.
APPLICANT: Watanabe, Colin K.
APPLICANT: Wood, William I.
TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
FILE OF INVENTION: ACIDS ENCODING THE SAME
FILE REFERENCE: P3230R1C1
CURRENT APPLICATION NUMBER: US/10/063,685
CURRENT FILING DATE: 2002-05-08
Prior Application removed - See Palm or File Wrapper
NUMBER OF SEQ ID NOS: 170
SEQ ID NO 33
LENGTH: 2773
TYPE: DNA
ORGANISM: Homo Sapien
-10-063-685-33

Query Match 7.98; Score 264.5; DB 14; Length 2773;
Best Local Similarity 43.48; Pred. No. 3e-06;
Matches 79; Conservative 11; Mismatches 85; Indels 7; Gaps 2;
449 GSRDCSTQTERGTESNKTAAVAPISVPAPVAAAAATAAATATATATTTTAA 503
2594 GTTTCATTTTGTCA--TGACAAATGTAGGAATGCTGAATTAATGTTTGAAGGATGAA 2651
504 VAVAAAAAPAAAAAPSPATATAATAAVSPAAAGQIPAAASVASAAVAPSAASAAVQVA 563
2652 AATAAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAA 2711
564 PAAPAPVPALVVPAPAAQASAPAOQAPTSAPAVAPTPTPTPAVAQAEVPSA 623
2712 AAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAA 2771
624 TG 625
2772 AG 2773

SULT 14
-10-369-493-3279
Sequence 3279, Application US/10369493
Publication No. US20030233675A1
GENERAL INFORMATION:
APPLICANT: Cao, Yongwei
APPLICANT: Hinkle, Gregory J.
APPLICANT: Slater, Steven C.
APPLICANT: Goldman, Barry S.
APPLICANT: Chen, Xianfeng
TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES

FILE REFERENCE: 38-10(52052)B
CURRENT APPLICATION NUMBER: US/10/369,493
CURRENT FILING DATE: 2003-02-28
PRIOR APPLICATION NUMBER: US 60/360,039
PRIOR FILING DATE: 2002-02-21
NUMBER OF SEQ ID NOS: 47374
SEQ ID NO 3279
LENGTH: 1965
TYPE: PRT
ORGANISM: Neurospora crassa
US-10-369-493-3279

Query Match 7.98; Score 263.5; DB 15; Length 1965;
Best Local Similarity 22.9%; Pred. No. 2.3e-06;
Matches 189; Conservative 110; Mismatches 280; Indels 245; Gaps 38;
26 QQWVEI-----LSDNNENLRQLEGCEYKVAELQKVEIQRVSEAYENLVKSSSKREA 79
1038 QETAETIKAREDDUTAQNKLLHQOLDGVTSQISALKOSRSAMDESAG-----TOISANDTA 1093
80 LE--KAMRNKL--EGEI-----RRMHDNFRDLRERLETANKQLAEKYEYGEDTK-- 126
1094 IEGLRELNNYLRRKEKILEVQFDLKVQEARN--LQQCLEYSQSOLDDEARLKLQDERRSQAD 1152
127 -----TISQIFAK-----NKESOREKEKLEAEIATARSTNEDQORRHEIRQOAL 170
1153 SARNSLTHKELDKLNLNLIRESNVTLRNENLRIOQLAMKVRKTIEDLENRIQPLEARI 1212
171 SN-----AAQKVVLEEB---LKKK-----QVYVDKVERMQQALVOLQAAACEKR 211
1213 SELELDKSPKEAEVQKQLEARDGLQKRISSILSKYQADPQVEVEQLKASVSESQAEREAF 1272
212 EQLEHRLRTEL---ERELESLR-----IQROGNCQPTNVSEYNA--AALME 253
1273 KQTEAALKEELKEAETNLETERTNWKNMRARLAEDPKTRFGNVK--TORNELAAEKQALQT 1331
254 LLREKEERILALEADMTKEQ-----KYLEENVMRHPALDAAATVAAQDDTTTVISHPNT 308
1332 TLDANERLSAVEKDLESARQDLANLQAOQNTLOEQHQAQAAAAAASAEQTPVAPAA 1391
309 S---YDTALEARIQ--KBESEILMANKRCIDMBGRIKTLHAQ-----IIRKD-AMIKVLQ 358
1392 APAEQDQLQOLDALRQELLESVKSKAALETE--LESKAEIATAIABRDEARAEVARL 1449
359 RSRKEPKETQLSCMRPAK--SLMSISNAGSGLLSHSSTLTGSP----- 401
1450 QSSTHTGDTTEMQDVEAPAPASAPAPAQEGTGLSDBERKALBERIAAAEAKAAEFKAK 1509
402 -MBEKRD-----DKSMKSGSLGILLGCD-----YRAY- 427
1510 DLEERADAIVKQSRKMKTKALNELAESKEAMEKQTDQERQKLAQYDLKLOQELAILKA 1569
428 -----VPSTPSPVP-----PSTPLLSAHSKTSGRD--CSTQTERG--- 460
1570 EQQSGVSGNGVPATPAKPPPTAQTQTPGAGTTPGLPLANLTDOOTRELISSNQVIMGIK 1629
461 -----TESNKTAAVAPISVPAP-----VAAAAATAAATATAATITTTMV----- 499
1630 SNVKNTAASKKVRRELEVVVVKAETQKIASKEQAALTEKKKSALRLNMLORQLKTEK 1689
500 -AAAPVAVAAAAAP-----AAAAASPATAAATAAVSPAAAGQIPAAASVA 545
1690 AKTAVVETAARETPOKPVWVWVWVWVWVWVWVWVWVWVWVWVWVWVWVWVWVWVW 1738
546 SAAVAPS-----AAAAAVQVAPAPAPVAPALVVPAPAAQASA----- 588
1739 QRARKSMTKXTNDGSDAAA VVVPK-AGEIEIKQIPTPTNTKAPAPAROGSIVGGQVVG 1797
589 -----PAQTQAP-TSAPAVAPTPTPTTTPAVAQAEVPSATGP 626
1798 NPFQPGANQSPVTSQFPANPFPGAAPOQQNAQPOQP--PATNP 1839

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

protein - protein search, using sw model

n on: June 16, 2004, 19:09:47 ; Search time 27.2311 seconds
(without alignments)
2384.384 Million cell updates/sec

US-09-332-063-2

Effect score: 3347

quence: 1 MPRAQSSASVQVPADPFA.....KTPIQILQGPDAEMVEYLI 675

oring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

arched: 283366 seqs, 96191526 residues

tal number of hits satisfying chosen parameters: 283366

nimum DB seq length: 0
ximum DB seq length: 2000000000

st-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

tabase : PIR 78:**

1: Piri:**
2: Piri:**
3: Piri:**
4: Piri:**

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

sult No.	Score	Query Match	Length	DB ID	Description
1	300	9.0	1087	1 QPMSH	neurofilament trip
2	292	8.7	507	2 T44768	antifreeze glycope
3	286.5	8.6	2346	2 T13829	Tpr homolog - frui
4	285.5	8.5	1072	1 A37221	neurofilament trip
5	280	8.4	3534	2 T42567	tegument protein 2
6	279	8.3	801	2 T29018	hypothetical prote
7	278	8.3	600	2 T29879	hypothetical prote
8	274.5	8.2	1299	2 T47182	hypothetical prote
9	263.5	7.9	352	2 A36128	regulatory protein
10	257.5	7.7	416	1 SKXLAG	dermal gland prote
11	252	7.5	858	2 S15762	neurofilament trip
12	250.5	7.5	352	2 G82990	arginate regulator
13	249	7.4	1020	1 QFHUH	neurofilament trip
14	248.5	7.4	1039	2 S18199	myosin heavy chain
15	245.5	7.3	2094	2 S33224	tpr protein - huma
16	243.5	7.3	1334	2 T50568	probable multi-dom
17	243	7.3	473	2 S50755	hypothetical prote
18	242	7.2	217	2 S29309	hypothetical prote
19	242	7.2	309	2 G83013	polynhydroxyalkanoa
20	242	7.2	915	2 S36327	clathrin assembly
21	240.5	7.2	865	2 A47282	calcium-binding pr
22	240.5	7.2	873	2 A47283	calphotin - frui
23	238.5	7.1	581	2 T22341	hypothetical prote
24	238.5	7.1	854	2 S02003	neurofilament trip
25	237	7.1	340	2 A35630	regulatory protein
26	236.5	7.1	1132	2 C75959	probable iron-sulf
27	234	7.0	901	2 A44825	phosphoprotein, sy
28	234	7.0	1110	2 T51116	NP-180 sea lamp
29	231.5	6.9	721	2 E70766	hypothetical prote

30	231	6.9	4776	2 B95206	cell wall surface
31	230	6.9	896	2 S36326	clathrin assembly
32	229.5	6.9	1952	2 T48614	hypothetical prote
33	227.5	6.8	1367	1 S48478	glucan 1,4-alpha-g
34	226.5	6.8	1017	2 PC4035	cell-cycle-depende
35	226	6.8	822	2 A38420	antifreeze glycopr
36	226	6.8	1794	2 T38459	hypothetical diver
37	225	6.7	1931	2 A59234	slow myosin heavy
38	224.5	6.7	1736	2 T00391	hypothetical prote
39	224	6.7	441	2 T24591	hypothetical prote
40	223.5	6.7	1938	1 JX0178	myosin heavy chain
41	223	6.7	1940	2 A59287	myosin heavy chain
42	222.5	6.6	215	2 S55925	probable arabinoga
43	222.5	6.6	1133	2 T22976	hypothetical prote
44	221.5	6.6	797	1 VGBX1	glycoprotein X pre
45	221	6.6	412	2 S07537	myosin heavy chain

ALIGNMENTS

RESULT 1

QPMSH
neurofilament triplet H protein - mouse
C:Species: Mus musculus (house mouse)
C:Date: 31-Mar-1992 #sequence revision 31-Mar-1992 #text_change 22-Jun-1999
C:Accession: J0368; A43778; S42616
R:Julien, J.P.; Cote, F.; Beaudet, L.; Sidky, M.; Flavell, D.; Grosveid, F.; Mushynski, Gene 68, 307-314, 1988
A:Title: Sequence and structure of the mouse gene coding for the largest neurofilament s
A:Reference number: J0368; MUID:89121513; PMID:3220257
A:Accession: J0368
A:Molecule type: DNA
A:Residues: 1-1087 <JUL>
A:Cross-references: GB:M23349; GB:M24496; NID:G200034; PIDN:AAA39813.1; PID:G387493
R:Shneidman, P.S.; Carden, M.J.; Lees, J.F.; Lazzarini, R.A.
Brain Res. Mol. Brain Res. 4, 217-231, 1988
A:Title: The structure of the largest murine neurofilament protein (NF-H) as revealed by
A:Reference number: A43778
A:Accession: A43778
A:Molecule type: mRNA
A:Residues: 'M', 1-132, 'QA', 134-199, 'R', 200-280, 'T', 282-491, 'G', 493-533, 'GEAKSP', 534-545,
A:Cross-references: GB:K35131; NID:G200021; PIDN:AAA39809.1; PID:G200022
R:Carden, M.J.
submitted to the EMBL Data Library, March 1994
A:Reference number: S42616
A:Accession: S42616
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-132, 'QA', 134-199, 'R', 200-280, 'T', 282-491, 'G', 493-533, 'GEAKSP', 534-545, 'R',
A:Cross-references: EMBL:Z31012; NID:G463249; PIDN:CAA83229.1; PID:G463250
C:Genetics:
A:Gene: nfh
A:Introns: 290/1; 356/3; 398/2
C:Superfamily: neurofilament triplet H protein
C:Keywords: coiled coil; cytoskeleton; heterotrimer; intermediate filament; nerve; phosph
F:1-97/Domain: amino-terminal <NTS>
F:98-408/Domain: rod #status predicted <ROD>
F:409-1087/Domain: carboxy-terminal <CTE>
F:519-886/Region: 6-residue repeats
F:520,526,532,538,544,550,556,562,568,574,580,586,592,598,604,610,616,622,628,634,640,64
73,885/Binding site: phosphate (Ser) (covalent) #status predicted
F:772/Binding site: phosphate (Thr) (covalent) #status predicted

Query Match 9.0%; Score 300; DB 1; Length 1087;
Best Local Similarity 21.6%; Pred. No. 5.1e-05;
Matches 169; Conservative 121; Mismatches 301; Indels 190; Gaps 30;

Oy 10 SYQVPADPFAIVSRAQOMVEILSDENRLRQLEGCYKVKVRLQKVFETIQVRSAYEN 69
Db 53 SVSSVSASPFRCAASS-----TDSLDTLSNGPCGVVAARSEKEQLQALNDRFAG 107

```
70 LVKSSSKREALEKAMRNKLEGERMHED-----FNRLRE-----RLETANKOL 113
108 YIDKVRQLEAHNS-----LEGEAAALQOQGRAANGELVEREVGRGAVILGARGOL 163
114 AEKEYEGESDTRKTI SOLFAKNKESQREKEKLEAEALATARNSTNEOORHIEIRDQALS-- 171
164 RLEQEHLLD-----IAHVQRQLDEEARQREEAFAAALAFAGAEAAARVELQKQALQ 219
172 -----NQAQKVKLEELK-----KKQVYVDKVKMOQALVOLQA-ACE 209
220 BECYLRHHQEEVEGELGLOQCGGAQAQAQAEARDALKCDVTSALRIQALGHAVQ 279
210 KREQLERHLRLTRLELESURIQORQNCQPTNVSEYNAALMELLREKEBERILALEA-- 267
280 SSLQSEEFVRDLSEAAKVNVTDMASQAQEBITEYR-----RQQAQNTTELKALK 331
268 ---DMTQEQKYLEENVVRHFDALDAATAAQAQEDT--TWISHSPNTSYDTALEARIQKEE 322
332 STKESLERQSELED---RH-----QADIASYQDAIQQLDSELRNKTWMAAQLR---E 379
323 EEILMANKRCLDME-----GRIKTLHAQI-IEKDAMI 353
380 YQDLINVMALDIEIAAYRKLLEGECECRIGFGPSPFLTEGLPKIPISITHIKVKSBEKI 439
354 KYLQORSRKE-----PSKTEQLSCHRPXAKSLMSISNAGSGLLSHSSTLTGSPIMEKRDOK 409
440 KVV-EKSEKETVIEGQTEIR-----VTEGVTEB--EDK 471
410 SWKSGSLG--ILLGDVAAEYVPTSPVPSTPLLSAHSKTSRDCSTQTERGTSNKTA 467
472 EAQOQGEBAEKEEELAATS--PABEAASPEKETSRVKEEAKSPGEAKSPGE 529
468 AVAPISVPVAAATAAAT-ATAATITTTMVAAPVAVAAAPAAAAAP----- 518
530 AKSPAEPKSPGEAKSPGEAKSPGEAKSPABPKSPABPKSPABPKSPABPKSPATVKSPGE 589
519 -----SPATAATAAAVSPAAGQIPAAASVASAAAVASAAAAPAAAAVQVAPA----- 565
590 AKSPAEPKSPGEAKSPGEAKSPGEAK-SPABAK-SPABAKSPABAKSPABPKSPABPKSP 647
566 APAPVPAPALVPVAPAAAQAAPAQOAPTSA-----PAVAPTPTPTTTPAVA--- 614
648 SEAKSPAEPKSPGEAKSPGEAKSPABPKSPABPKSPABPKSPABPKSPABPKSPABPKSP 707
615 -----QAEVPA---SPATGCPGPHRLISFLTCNPDKTDGPFVHFHNTLERTPTIQLQ 664
708 AEVKSPPGEAKSPABPKSPABPKSPABPKSPABPKSPABPKSPABPKSPABPKSPABPKSP 767
665 E 665
768 E 768

SULT 2
4768
tiffreeze glycopeptide AFGP polyprotein precursor [imported] - Boreogadus saida
Species: Boreogadus saida
Date: 21-Jan-2000 #sequence_revision 21-Jan-2000 #text_change 21-Jul-2000
Accession: T44768
Chen, L.; DeVries, A.L.; Cheng, C.H.C.
oc. Natl. Acad. Sci. U.S.A. 94, 3817-3822, 1997
Title: Convergent evolution of antifreeze glycopeptides in Antarctic notothenioid fish
Reference number: 222834; MUID: 97268653; PMID: 9108061
Accession: T44768
Status: preliminary; translated from GB/EMBL/DBJ
Molecule type: DNA
Residues: 1-507 <CH>
Cross-references: EMBL:U43200; NID:G2078482; PIDN:AAC60129.1; PID:G2078483
Genetics:
  Introns: 1/3

Query Match      8.7%; Score 292; DB 2; Length 507;
Best Local Similarity 35.0%; Pred. No. 4.6e-05;
```

```
Matches 86; Conservative 19; Mismatches 91; Indels 50; Gaps 6;
Qy 433 SPVPPSTLLSAHSKTSRDCSTQTERGTSNKTAAPV--ISVPAPVAAATA----- 484
Db 66 TPATAATPATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 125
Qy 485 -----AAATATAATITTTMVAAPVAVAAAPAAAAAP-SPATAAAT 526
Db 126 ETPARAATPATAATPATAATAATAATAATAATAATAATAATAATAATAATAATA 185
Qy 527 AAASVSPAAGQIPAAASVASAAAVASAAAAPAAAAVQVAPA-PAPVPAPALVPV 585
Db 186 ARAATPATAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 245
Qy 586 ASAP-----AQOAPTSPAPAVAPTPTPTTTPAV-----AQAEVPA 622
Db 246 ATAPTATPAAAAAATPATAATLATAATAATPATAATPATAATAATAATAATAAT 305
Qy 623 ATGPGP 628
Db 306 ATAATP 311

RESULT 3
Tl3829
Tpr homolog - fruit fly (Drosophila melanogaster)
C/Species: Drosophila melanogaster
C/Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 17-Nov-2000
C/Accession: Tl3829
R/Zimowska, G.; Aris, J.P.; Paddy, M.R.
J. Cell Sci. 110, 927-944, 1997
A/Title: A Drosophila Tpr protein homolog is localized both in the extrachromosomal chan
A/Reference number: 217786; MUID: 97296455; PMID: 9152019
A/Status: preliminary; translated from GB/EMBL/DBJ
A/Molecule type: mRNA
A/Residues: 1-2346 <ZIM>
A/Cross-references: EMBL:U91980; NID:G1923273; PID:G1923274; PIDN:AAC47506.1
C/Genetics:
A/Cross-references: FlyBase:FBgn0013756
A/Map position: 2R

Query Match      8.6%; Score 286.5; DB 2; Length 2346;
Best Local Similarity 22.5%; Pred. No. 0.0041;
Matches 157; Conservative 124; Mismatches 259; Indels 157; Gaps 32;
Qy 14 VPADPPALVSPAAQVVEILSDENRNLRSQLECYKVA---RLQKVETEI-----Q 61
Db 1262 VSANKHEEVLRKIETLNATDSNRILREARNALTVABELTRISSVEKELPQCSNKE 1321
Qy 62 RVSEAYE-NLVKSSSKREALEKAMRNK--LEGEIRMHDFNRDLRRLRLETANKQLAEKY 118
Db 1322 LTSKIEINVENTSLRTEAIKWRQRANALVEKSNRNPFEFKLQAEERHLAKILLTAEKEL 1381
Qy 119 EGSEDYRKTI-----SOLFANKESQ---REKELEAELATARNSTNEOORHI-BIRDQ 168
Db 1382 NKQSDSLTVLKQRMTEIPMLNKKQOILDEARKQVDEFTMLKQNTQTQDIMEKKR 1441
Qy 169 ALSNAQAKVVKLEELK-KKQVYVDKVKMOQALVOLQAACSKREQLERHLR-----TL 222
Db 1442 LLQK-BEELKANELETKDKTADTKETK---LQLRLAKRYKDFYIGLOSQGGTSES 1496
Qy 223 ERELESURIQORQNCQPTNVSEYNAALMELLREKEERI-----LALEADMTKW 272
Db 1497 AAELKVRSELEEVNQ-----LRLAKBHEKITKECDVKKRTPETUTSAI 1544
Qy 273 EOKY---LEENVVRHFDALDAATAAQAQAEARDLTWISHSPNTSYDTALEARIQKEEELMAN 329
Db 1545 ROEKAKLKDILVV---DLTVARTDLVNQETTF--GTSKYDETI-ARLEKELQENIVAN 1598
Qy 330 KRCLDMEGRI-----KTLHAQIIEKDAMIKVLOORSRKEPKTEQLSCMPAKSLAISNA 385
Db 1599 K---DINQRLTRENESLHMRNQ-----LTRLQSGQOSTKPTSTS-SVA 1637
```


510 ANAKSSLE--KAKHLQNEQEDWMDLEKANSAAASLDKKQSGPKIINDWKQKYEES 566
366 KTEQSCMRPAKS---LMSISNAGSGLLSHSSTLTGSPIMEEKRDDKS 410
567 QAELEASQKARSSTELFKNAVEETLDHLETL-----KXENXN 607

protein - human
alternate names: kinase-related transforming protein (tpr-met); protein with promoter
species: Homo sapiens (man)
date: 22-Nov-1993 #sequence revision 26-May-1995 #text_change 11-Jan-2002
citation: S33124; S23740; S00928; G01185
itcheil; P.J.; Cooper, C.S.
cogene 7, 2323-2333, 1992
title: The human tpr gene encodes a protein of 2094 amino acids that has extensive coiled-coil structure
reference number: S33124; PMID:93064711; PMID:1437155
accession: S33124
status: preliminary; nucleic acid sequence not shown; translation not shown
clecule type: mRNA
esidues: 1-2094 <MIT>
ross-references: EMBL:X66397; NID:9633225
note: the nucleotide sequence was submitted to the EMBL Data Library, October 1991
itcheil; P.J.; Cooper, C.S.
cogene 7, 383-388, 1992
title: Nucleotide sequence analysis of human tpr cDNA clones.
reference number: S23740; PMID:92195670; PMID:1549355
accession: S23740
status: preliminary
clecule type: mRNA
esidues: 1-725, 'L' <M12>
ross-references: EMBL:X63105; NID:g37257; PIDN:CAA44819.1; PID:g37258
ing, H.W.S.; Tempest, P.R.; Merrifield, K.R.; Rance, A.J.
cogene 2, 617-619, 1988
title: Tpr homologues activate met and raf.
reference number: S00928; PMID:88262257; PMID:3387099
accession: S00928
clecule type: mRNA
esidues: 1-31, 'R', 33-142 <KIN>
ross-references: EMBL:Y00672; NID:g37255; PIDN:CAA68681.1; PID:g37256
reco, A.
note: submitted to the EMBL Data Library, December 1995
reference number: H00592
accession: G01185
status: translated from GB/EMBL/DBJ
clecule type: DNA
esidues: 144-228 <GRE>
ross-references: EMBL:X94208; NID:g1296797; PIDN:CAA63904.1; PID:g1296798
natics:
ne: GDB:TPR
ross-references: GDB:128821; OMIM:189940
p position: 1q25-1q25
itrons: 177/3

ery Match 7.3%; Score 245.5; DB 2; Length 2094;
st Local Similarity 20.3%; Pred. No. 0.016;
tches 153; Conservative 127; Mismatches 274; Indels 201; Gaps 31;

24 RAQQVVEILSDENRLRQEGCEKVK---ARLQKVEIQRVSEAVENLVKSSK---- 76
1223 RYRQVVELLELELEDSLNAEKQVYVAKTAQAHEELMKKTETM-NVVMETNOLRE 1281

77 -REALEKAMR-----NKLEGEIRMDNFNRLDER--LETANKQLAEKYE----- 119
1282 EKERLEQDLOQMAQKVRKLELDILFLOEANAELSEKGMQLQAEKLLLEEDVKWKARNQH 1341

120 -----GSEDRKTVISQFAKNKESORBEK---KLEAEIATARSTNEDQRHIERDQ 168
1342 LVSQQKDPDTEYERKLLSEKEVHTKRIQOLTTEEIGRLKAEIARNA----- 1387

169 ALSNAQAKVVKLEELKKKKVYVDKVEK-MQQAUVQL-QAAACEKREOLEHRLRRLERELE 227

Search completed: June 16, 2004, 19:15:39
Job time : 31.2311 secs

Db 1388 SLTNNQNLIOSLKEDLNKVRTEKTIQKDLAKIIDIQEKVKTIQVK-KIGERYKTOYE 1446
QY 228 SLRIO-----ORQNCOPNNV-----EYNAALMELLRE 257
Db 1447 ELKAOQKVMETSQSSGDHQBQHVSVQEQELKETLQAEYKSKLSQVENLOKTLSE 1506
QY 258 KB-----BRILALGDMTKWEQKYLENVNRHFPALDAAATVAAQORD-TTVISHSPNYS 309
Db 1507 KETEARNLOEQTVLOQSELRLQD-LQDRITTOBQLOQOIITEKBEKTKAIVAAKXIA 1565
QY 310 YDTALEARIQKEEIEILMAN-----KRCIDM-----EGRIKTLHAQIIEKIDAMIK 354
Db 1566 HLAGVKDQLTKEBELKQNGALDQKDELVRITALKSOYEGRISRLERELREHQB--R 1623
QY 355 VLQOR-----SRKEPSKTEQLSCMRPAKSIAMISNAGSGLLSHSSTLTGSPIMEEKRDD 408
Db 1624 HLEQRDBQPSNKNVPEQCRQI-----TLKTTFA----- 1652
QY 409 KSMKSGILGIDYRAEVVPSPTSPVPPSTPLLSAHSKTSGRDCSTQTERGTSNKTAA 468
Db 1653 ---SGERGIASTSD-----PPTANIKP--TPVSTPSKV-----TAAAMAGNKTPTAS 1696
QY 469 VAPISVPAPVAAAATAATATATATATTTTWAAAAPVAVAAAAAPAAAAAPSPA-TAAATA 527
Db 1697 IRPMVTPTATVNTPT-----TPTATVMTTQVESQEAQMOSEGVPVHVFGTSGSVRSTS 1752
QY 528 AAVSPAAAGQIPAAASVAGAAAAPVAPAAAAAVQVAPAPAPVAPALVPVAPAAAAQAS 587
Db 1753 PNVQPSISQPI-----LTVOQQTQATFVQ-----PTQOSHPIEPANQELSSNIV 1798
QY 588 APAQT---QAPTSAPAVATPATPTTPAV-----AQAEVPASPATGPGPHR 630
Db 1799 EVVQSSPVPERSTSTAVFGTVSATPSSSLPKTRBEEEDSTIENSQVSDDTVEMPLEPKK 1858
QY 631 LSTPSILTCNPDKTDGVPFHSNLTLEKTPITQILQOE 665
Db 1859 LK---SVTPVGTETEEVMAEESDGEVETQVYNQD 1869

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

protein - protein search, using sw model

on: June 16, 2004, 19:12:42 ; Search time 16.5037 Seconds
(without alignments)
2129.669 Million cell updates/sec

le: US-09-332-063-2
fect score: 3347
puence: 1 MPRAQPSASVQVPADPFA.....KTPIQILGQEPDAEMVEYLI 675

ring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

arched: 141681 seqs, 52070155 residues

al number of hits satisfying chosen parameters: 141681

imum DB seq length: 0
imum DB seq length: 2000000000
t-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

abase : SwissProt_42.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

sult No.	Score	Query Match	Length	DB	ID	Description
1	300	9.0	1087	1	NFH_MOUSE	P19246 mus musculus
2	292.5	8.7	518	1	TPM4_DROME	P49455 drosophila
3	257.5	7.7	439	1	XP2_XENLA	P17437 xenopus lae
4	252	7.5	857	1	NFM_CHICK	P16053 gallus gall
5	250.5	7.5	352	1	ALGE_PSEAE	P15276 pseudomonas
6	246	7.3	1026	1	NFH_HUMAN	P12036 homo sapien
7	245.5	7.3	2349	1	TPR_HUMAN	P12270 homo sapien
8	242.5	7.2	1102	1	MYSC_CHICK	P05140 rattus norv
9	242	7.2	915	1	A180_RAT	C07788 bos taurus
10	241.5	7.2	831	1	NFM_BOVIN	P16884 rattus norv
11	241.5	7.2	865	1	CPN_DROME	Q02910 drosophila
12	240.5	7.2	1157	1	BECI_YEAST	P47068 saccharomyc
13	239.5	7.1	790	1	BEAL_MOUSE	Q8B166 mus musculus
14	239	7.0	901	1	A180_MOUSE	Q61548 mus musculus
15	234	6.9	721	1	YK82_MYCTU	Q10690 mycobacteri
16	231.5	6.9	1217	1	ITN1_RAT	Q9WVE9 rattus norv
17	230.5	6.9	1217	1	MYSS_CYPCA	Q90339 cyprinus ca
18	229.5	6.9	1935	1	ENAH_MOUSE	Q03173 mus musculus
19	229	6.8	1802	1	AMPH_YEAST	P08640 saccharomyc
20	227.5	6.8	1367	1	CENF_MOUSE	P49454 homo sapien
21	226.5	6.8	3210	1	ANP_NOTCO	P24856 notothenia
22	226	6.8	790	1	YDC9_SCHPO	Q10172 schizosacch
23	226	6.8	1794	1	A180_HUMAN	Q06041 homo sapien
24	224.5	6.7	907	1	MYSS_CHICK	P13538 gallus gall
25	223.5	6.7	1938	1	3BP1_HUMAN	Q9Y313 homo sapien
26	223	6.7	622	1	VGLX_HSVB	P28968 equine herp
27	221.5	6.6	797	1	MYH9_CHICK	P14105 gallus gall
28	221	6.6	1959	1	ITN1_MOUSE	Q9Z074 mus musculus
29	218.5	6.5	1714	1	ITN1_HUMAN	Q15811 homo sapien
30	218	6.5	1721	1	MYSU_RABIT	Q99105 oryctolagus
31	217.5	6.5	501	1	MYHA_RAT	Q9J1K0 rattus norv
32	217.5	6.5	1976	1	MYA_RAT	P24733 aequipecten
33	217	6.5	1938	1	MYA_RAT	

34 216.5 6.5 1976 1 MYHA_BOVIN Q27991 bos taurus
35 216 6.5 1976 1 MYHA_HUMAN P35580 homo sapien
36 214.5 6.4 1238 1 SBCC_RHOCA O68032 rhodobacter
37 214.5 6.4 2442 1 CEP2_HUMAN Q9BV73 homo sapien
38 214 6.4 1590 1 C190_DROME Q9VJES drosophila
39 213.5 6.4 1546 1 NUP1_DROME Q9VDV3 drosophila
40 213.5 6.4 1935 1 MYH7_HUMAN P12883 homo sapien
41 213.5 6.4 1940 1 MYH7_RAT P12847 rattus norv
42 213 6.4 1972 1 MYHB_MOUSE O08638 mus musculus
43 212.5 6.3 1935 1 MYH7_PIG P79293 sus scrofa
44 212.5 6.3 1938 1 MYSD_CAEEL P02567 caenorhabdi
45 212.5 6.3 2517 1 NCR2_HUMAN Q9Y618 h nuclear r

ALIGNMENTS

RESULT 1
NFH_MOUSE
ID NFH_MOUSE STANDARD; PRT; 1087 AA.
AC P19246; Q61959;
DT 01-NOV-1990 (Rel. 16, Created)
DT 01-AUG-1992 (Rel. 23, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Neurofilament triplet H protein (200 kDa neurofilament protein)
DE (Neurofilament heavy polypeptide) (NF-H).
GN NEFH OR NFH.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=89121513; PubMed=3220257;
RA Julien J.-P., Cote F., Beaudet L., Sidky M., Flavell D., Grosveld F.,
RA Mushynski W.;
RT "Sequence and structure of the mouse gene coding for the largest
RT neurofilament subunit";
RL Gene 68:307-314(1988).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=89089138; PubMed=3145094;
RA Shneiderman P.S., Carden M.J., Lees J.P., Lazzarini R.A.;
RT "The structure of the largest murine neurofilament protein (NF-H) as
RT revealed by cDNA and genomic sequences";
RL Brain Res. 464:217-231(1988).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=Swiss Webster; TISSUE=Brain;
RA Carden M.J.;
RL Submitted (MAR-1994) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: Neurofilaments usually contain three intermediate
CC filament proteins: L, M, and H which are involved in the
CC maintenance of neuronal caliber. NF-H has an important function in
CC mature axons that is not subserved by the two smaller NF proteins.
CC -!- PTM: There are a number of repeats of the tripeptide K-S-P. NFH is
CC phosphorylated on a number of the serines in this motif. It is
CC thought that phosphorylation of NFH results in the formation of
CC interfilament cross bridges that are important in the maintenance
CC of axonal caliber.
CC -!- PTM: Phosphorylation seems to play a major role in the functioning
CC of the larger neurofilament polypeptides (NF-M and NF-H), the
CC levels of phosphorylation being altered developmentally and
CC coincident with a change in the neurofilament function.
CC -!- SIMILARITY: Belongs to the intermediate filament family.
CC -!- CAUTION: Ref.2 sequence differs from that shown in positions 534
CC to 716 and is shorter due to frameshifts.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial

```

ery Match          9.0%; Score 300; DB 1; Length 1087;
st Local Similarity 21.6%; Pred No. 5,1e-05;
atches 169; Conservative 121; Mismatches 301; Indels 190; Gaps 30;

10 SYQVPADFAIVSRAQQMVEILSDENNRLRQELGECYKVARLQKVETIQRVSEAYEN 69
      :|||:
53 SVSSVSASPSRFRGAASS-----TDSLDTLSNGPEGCVVAAVAARSEKEQLQALNDRFAG 107
      :|||:
70 LVKSSSKREALKAMRNKLEGEIRMHM-----FNRLRE-----RLETANKQL 113
      :|||:
108 YDKVTRQLEAHNRS-----LEGEAALRQOKGAAMGELYREVRMEGAVLRGAEGL 163
      :|||:
114 AKKEYGSEDTRKTIQSOLFAPKYESQRKEKLEAELATARNSTEDQRHTEIRDOALS-- 171
      :|||:
164 RLEQEHLLSD-----IAHVRQRLDEEAROREEAEAAARALAPAEAAEARVELQKKAQL 219
      :|||:
172 -----NAQAKWKVLEELK-----KKQVYVDKVKQKQALVOLQA-ACE 209
      :|||:
220 BECCYLRRHQBEVVELLQIOGCGAAQAQAQAEARDALKCDVTSALREIPALQEGHAVQ 279
      :|||:
210 KREQLBHRUTRLRLEHESLRIQQPGNQCPNTVSEYNAAMLELKEEBRIUALBA-- 267
      :|||:
280 SSLQSENEFRVRLDRLSEAAKVNTDAMSAOBEITEYR-----RQLQARTTELEALK 331
      :|||:
268 ---DMTKQKQKYLEENVMGHFALDAAATVAAORDT--TWISHSPTSVDYALAEARIQKE 332
      :|||:
332 STKESLERQSELED---RH-----QADIASYQDAIQQLDSELNTRKWBMAAQLR---E 379
      :|||:
323 EETIMANKVCLDME-----GRIKTLHAQI-LEKDAMI 353
      :|||:
380 YQDLLNVKQALDIETAAAYRKLLEGECRIGFSPSPFTEGLPKIPISITHIKVSEMI 439
      :|||:
354 KVLQQRSRKE-----PSKTRBOLSCMRPAKSLMSISNAGSGLLSHSSTLTGSPIMEKRD 409
      :|||:

```

440 KVV-EKSEKTVIVEGQTEER- - - - -VTEGVTEB--EDX 471

410 SWKSLG--ILLGGDYRAYEYVPTSPVPVPSTPLLSAHSKTSGRDCSTQTERGTGTSNKTA 467

472 EAQQQGEGBEABEGEBKBEELAAATS--PPAEAAASPEKETKSRVKBERAKSPGEAKSPGE 529

468 AVAPISVPAPVAAAATAAAT-ATAATITTTVAAAAPVAVAAAAAPAAAAAP- - - - - 518

530 AKSPAEAKSPGGEAKSPGGEAKSPGEAKSPAEPKSPAEKSPAEKSPAEKSPAEKSPAEKSPGE 589

519 - - - - -SPATAAAATAAAVSPAAAGQIPAAASVASAAAAVAPSAAAAAVQVAPA- - - - - 565

590 AKSPSEAKSPAEKSPAEKSPAEPKSPAEKSPAEPKSPAEKSPAEPKSPAEKSPAEKSPGEAKSP 647

566 APAPVPAPALVPVAPAPAAQAASAPACTQAPTSA- - - - -PAVAPTAPPTTAPVA- - - 614

648 SEAKSPAEKSPAEPKSPAEPKSPAEPKSPAEPKSPAEPKSPAEPKSPAEPKSPAEPKSPAEPKSP 707

615 - - - - -QAEVPA--SPATGPGPHRLSIPSLTNCNPDKTDGVPFHSTLERTPIQILQO 664

708 AEVKSPEAKSPAAYKSPAEPKSPAAYKSPGGEAKSPGGEAKSPAEPKSPAEPKSPAEPKSP 767

665 E 665

768 E 768

RESULT 2

TPM4_DROME

ID TPM4_DROME STANDARD; PRT; 518 AA.

AC P49455; P49456; Q24425; Q24426;

01-FEB-1996 (Rel. 33, Created)

01-FEB-1996 (Rel. 33, Last sequence update)

10-OCT-2003 (Rel. 42, Last annotation update)

DE Tropomyosin I, isoforms 33/34 (Tropomyosin II).

GN TMI OR TMI1.

OS Drosophila melanogaster {Fruit fly}.

OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;

OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;

OC Ephydroidea; Drosophilidae; Drosophila.

OX NCBI_TaxID=7227;

RP [1]

RN SEQUENCE FROM N.A. [ISOFORMS 33 AND 34].

RP TISSUE=Embryo, and Pupae;

RC MEDLINE=89127197; PubMed=2851721;

RC Hanks P.D., Storti R.V.;

RA "The Drosophila melanogaster tropomyosin II gene produces multiple

RT proteins by use of alternative tissue-specific promoters and

RT alternative splicing";

RL Mol. Cell. Biol. 8:3591-3602(1988).

[2]

RN SEQUENCE FROM N.A. [ISOFORMS 33 AND 34].

RP STRAIN=Oregon-R; TISSUE=Pupae;

RC MEDLINE=87064486; PubMed=3097506;

RC Karlik C.C., Fyrberg E.A.;

RA "Two Drosophila melanogaster tropomyosin genes: structural and

RT functional aspects.";

RL Mol. Cell. Biol. 6:1965-1973 (1986).

CC -1- ALTERNATIVE PRODUCTS:

CC Event=Alternative splicing; Named isoforms=5;

CC Comment=Additional isoforms seem to exist;

CC Name=33; Synonyms=9C;

CC IsoId=P49455-1; Sequence=Displayed;

CC Name=Muscle; Synonyms=9D;

CC IsoId=P06754-1; Sequence=External;

CC Name=Non-muscle; Synonyms=Cytoskeletal;

CC IsoId=P06754-2; Sequence=External;

CC Name=9A;

CC IsoId=P06754-3; Sequence=External;

CC Name=34; Synonyms=9B;

CC IsoId=P49455-2; Sequence=VSP_006623, VSP_006624, VSP_006625;

CC -1- TISSUE SPECIFICITY: Both isoforms are only expressed in indirect

CC flight muscles.

-!- DEVELOPMENTAL STAGE: Both isoforms are expressed during pupal and adult stages.
 -!- DOMAIN: The molecule is in a coiled coil structure. The sequence exhibits a prominent seven-residues periodicity.
 -!- SIMILARITY: Belongs to the tropomyosin family.

THIS SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

EMBL; X76208; CAA53800.1; .
 EMBL; X76208; CAA53801.1; .
 EMBL; K02620; AAA28967.1; ALT SEQ.
 EMBL; K02620; AAA28967.1; JOINED.
 EMBL; L00355; AAA28967.1; JOINED.
 EMBL; L00356; AAA28967.1; JOINED.
 EMBL; L00357; AAA28967.1; JOINED.
 EMBL; L00358; AAA28967.1; JOINED.
 EMBL; L00359; AAA28967.1; JOINED.
 EMBL; L00360; AAA28967.1; JOINED.
 EMBL; L00361; AAA28967.1; JOINED.
 EMBL; L00362; AAA28967.1; JOINED.
 EMBL; M12840; AAA28967.1; JOINED.
 EMBL; K02621; AAA28968.1; .
 EMBL; M12840; AAA28968.1; JOINED.
 EMBL; L00355; AAA28968.1; JOINED.
 EMBL; L00356; AAA28968.1; JOINED.
 EMBL; L00357; AAA28968.1; JOINED.
 EMBL; L00358; AAA28968.1; JOINED.
 EMBL; L00359; AAA28968.1; JOINED.
 EMBL; L00360; AAA28968.1; JOINED.
 EMBL; L00361; AAA28968.1; JOINED.
 EMBL; L00362; AAA28968.1; JOINED.
 FlyBase; FBgn0003721; Tm1.
 GO; GO:004541; P:pole plasm oskar mRNA localization; IMP.
 InterPro; IPR000533; Tropomyosin.
 Pfam; PF00261; Tropomyosin.1.
 PRINTS; P800194; TROPOMYOSIN.
 PROSITE; PS00326; TROPOMYOSIN.1.
 Muscle protein; Cytoskeleton; Actin-binding; Coiled coil;
 Alternative splicing; Multigene family.
 DOMAIN 14 267
 COILED COIL (POTENTIAL).
 ALA/PRO-RICH.
 DOMAIN 287 518
 VARSPLIC 259 293
 IKSIEHVALNGDQDWTFTVEMGMCPFFYERY (in isoform 34).
 /Ftld=VSP_006623.
 VARSPLIC 300 367
 KLPTPTPEELAMEERAPAAAAAARAAAAAARAAAAAAGEAG
 ADGAPAPGSEKAPAKEPTPKKEPTTP -> ELTEEEKAL
 EAAIAAKARAEELALGEGAGAEAGGAPAGAPGAPG
 AATGVEAPAPPERIPT (in isoform 34).
 /Ftld=VSP_006624.
 VARSPLIC 391 518
 KNTPEPPGSEPPFVPAEGEAPAPAEAGAPPAEGAPPAE
 GAVPPADGAPPAEGAPPAEGAPPAEGAPPAEGAPPAE
 EAPPA -> RNAPGDFAPPAEAPPAEGAPPAEGAPPAE
 APPEPPAEGAPPAEGAPPAEGAPPAEGAPPAE
 /Ftld=VSP_006625.
 CONFLICT 106 114
 LGSAATKLS -> SASAIQLAA (IN REF. 2).
 A -> S (IN REF. 2).
 CONFLICT 119 119
 A -> S (IN REF. 2).
 CONFLICT 183 183
 A -> ANVEADLERABERA (IN REF. 2).
 CONFLICT 199 199
 P -> L (IN REF. 2).
 CONFLICT 503 503
 P -> A (IN REF. 2; AAA28968).
 SEQUENCE 518 AA; 54558 MW; 153D0872CF9DB6EA CRC64;
 Query Match 8.7%; Score 292.5; DB 1; Length 518;
 Best Local Similarity 24.1%; Pred.No. 4.6e-05;
 Matches 156; Conservative 76; Mismatches 246; Indels 169; Gaps 26;

12 QPVADPPIVSRQQMVEILSDENRNLRLQELGCEYKVARLQK- - -VETETIQRVSEAY 67

Db 9 QAMKVDKDGALERAL- - - - -VCEQEDANDTRAKEAEHARQLQKCIQTVENELDTQTEA- 63
 Qy 58 ENLVKSSSKREALEKAMRNKLEGEIRRMHDFNDRERLETANKQLAEKVEGESEDTKTK 127
 Db 64 - - - - -LTVTKLEKRNALQN-ABSEVAAL- - - - -NRI- - - - -QLLEEDLERSEERLGS 108
 Qy 128 ISQLFKNKESQREKLEALATARNSTNEDQRHRIEIRDOALSNAQKVVKLEELKKK 187
 Db 109 ATAKLSEASQAADSESR- - - - -ARKILENRALADBERMDALENQLKEARFLAEADKK 161
 Qy 188 QYVYDKVKQQAALVOLQAACEKREQLERLRLTELERE- - - - -LESRLQOQGNQCP 240
 Db 162 - - - - -YDEVARKLAWADLERAEERAEQGENKI-VELSEELRVVGNLKSLEVSSEKAN-QR 217
 Qy 241 TNVSEYNAALMELLREKERRILALEADMTKWKQKLEENVMRHFALDAATAATVAARDTT 300
 Db 218 EEEYKNQIKTLNTRLEKEAR- - - - -AEFAERSVQK- - - - - 248
 Qy 301 VISHPNTSYDTALEARIQKEEBEILMANVR-CLDMEGRIKTLHAQIIEKDMIKVLQ- - 357
 Db 249 - - - - -LQKEVDRLDDDLIVEKERYCMIGD- - - - -SLDEAFVD- - - - -LIKGLEPF 288
 Qy 358 - - - - -QRSRKEPS- - - - -KTEQLSCMRPAKSLMSISNAGSLLSHSSTLT- - - - -GSPIM- 402
 Db 289 WNPKNPPTPKLPTPTPEELAMEERAPAAAAAARAAAAAARAAAAAGAGADGAPAP 348
 Qy 403 -EKRRDKGKWSGLILLGDYRAEYVPTSPVPSTPLLSAHSKTSKSDCSTQTRGT 461
 Db 349 CEKAPAKE- - - - -PTPKKSTPTPPPPPPPEY- - - - -SIDLPPGEAEPV 388
 Qy 462 ESNKTAAVAPISVPVAAAAATAATTAATITTTVAAAAAPVAAAAAPAAAAAPSPA 521
 Db 389 YKKNYEPGSEPEPVAA- - - - -EGEAPAAAGAAPPA 423
 Qy 522 TAAATAAASVFAAGQIPAAASVASAAVAPAAAAVAPAAAAVAPAAVAPVAPVAPVAP 581
 Db 424 EGAA- - - - -PPAEGAVPPADGAPPAEGAPPAEGAPPAEGAPPAEGAPPAEGAP 469
 Qy 582 AAAASAPACTQAPTSAPAVAPTPTPTTAVQAQAEVPAASPATGPGP 628
 Db 470 APDAAAPAAEAPAAEAPAAEAPAAEAPAAEAPAAEAPAAEAPAAEAPAAEAPAAEAP 516

RESULT 3

XP2_XENLA STANDARD; PRT; 439 AA.
 AC P17437; Q08944;
 DT 01-AUG-1990 (Rel. 15, Created)
 DT 01-FEB-1995 (Rel. 31, Last sequence update)
 DT 15-MAR-2004 (Rel. 43, Last annotation update)
 DE Skin secretory protein xp2 precursor (APEG protein).
 OS Xenopus laevis (African clawed frog).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipidoidea; Pipidae;
 OC Xenopodidae; Xenopus.
 RN NCBI_TaxID=8355;
 RN [1]
 RP SEQUENCE OF 1-25 AND 344-439 FROM N.A. (ISOFORM 2).
 RX MEDLINE=92332564; PubMed=1629230;
 RA Hauser P., Roeben C., Hoffmann W.;
 RT "xp2, a new member of the P-domain peptide family of potential growth
 RT factors, is synthesized in Xenopus laevis skin."
 RJ J. Biol. Chem. 267:14451-14455(1992).
 RN [2]
 RP SEQUENCE OF 3-439 FROM N.A. (ISOFORM 1).
 RC TISSUE=Skin;
 RX MEDLINE=90127399; PubMed=2298293;
 RA Gmachl M., Berger H., Thalhammer J., Kreil G.;
 RT "Dermal glands of Xenopus laevis contain a polypeptide with a highly
 RT repetitive amino acid sequence."
 RJ FEBS Lett. 260:145-148(1990).
 CC -!- FUNCTION: May act as a growth factor in the germinal layer of the

epidermis. May also be involved in growth of regenerating glands and in protection of the skin from the external environment.

-!- SUBCELLULAR LOCATION: Secreted.
 -!- ALTERNATIVE PRODUCTS:
 Event=Alternative splicing; Named isoforms=2;
 Name=1; Synonyms=AP5EG;
 IsoId=P17437-1; Sequence=Displayed;
 Name=2; Synonyms=XP2;
 IsoId=P17437-2; Sequence=VSP_004652;
 -!- TISSUE SPECIFICITY: Skin.
 -!- SIMILARITY: Contains 2 P-type (trefoil) domains.
 -!- CAUTION: Ref.2 sequence differs from that shown from position 392 onward and is shorter (418 AA) due to a frameshift.

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

EMBL; M90095; AAA50001.1; -;
 EMBL; X51394; CAA35759.1; ALT_FRAME.
 PIR; A37331; A37331.
 PIR; S07498; SKXLAG.
 HSSP; P04155; 1PS2.
 InterPro; IPR000519; P_trefoil.
 Pfam; PF00088; trefoil; 2.
 PRINTS; PR00680; PTREFOIL.
 SMART; SM00018; PD; 2.
 PROSITE; PS00025; P_TREFOIL; 2.
 Signal; Growth factor; Alternative splicing; Repeat;
 Pyroglutamate carboxylic acid.
 SIGNAL 1 22 POTENTIAL.
 CHAIN 23 439 SKIN SECRETORY PROTEIN XP2.
 MOD RES 23 23 PYROLIDONE CARBOXYLIC ACID (PROBABLE).
 DOMAIN 26 343 33 X REPEATS OF G-[GE]-[AP] (2,4)-A-E.
 DOMAIN 350 391 P-TYPE 1.
 DOMAIN 397 438 BY SIMILARITY.
 DISULFID 351 377 BY SIMILARITY.
 DISULFID 361 376 BY SIMILARITY.
 DISULFID 371 388 BY SIMILARITY.
 DISULFID 398 424 BY SIMILARITY.
 DISULFID 408 423 BY SIMILARITY.
 DISULFID 418 435 BY SIMILARITY.
 VARSPLIC 26 343 Missing (in isoform 2).
 CONFLICT 3 3 H->S (IN REF. 2).
 CONFLICT 18 18 C->W (IN REF. 2).
 SEQUENCE 439 AA; 41173 MW; 38C4A4B57CBAE778 CRC64;

Query Match 7.78; Score 257.5; DB 1; Length 439;
 Best Local Similarity 34.2%; Pred.No. 0.00088;
 Matches 83; Conservative 16; Mismatches 113; Indels 31; Gaps 8;
 432 PVPPTPLLSAHSKTSRDCSTQVETGRTESNKTAAVAPISVPAPVAAAATAAATA 491
 119 PAPAPAGEAPAPAPAGEAPAPAPAE-GEAPAPAGEAPAPAPAEVEAPAPAPAGEAP 177
 492 ATITTTTVAAPVAV-----AAAAPAAAAPSPATAATAAAVSPA-AACQIPA----- 540
 178 APAPAGEAPAPAPAGEAPAPAPAGEAPAPAPAGEAPAPAPAGEAPAPAPAGE 237
 541 --AASVAAAAPSAAPAAAQVAPAPAPAPAPAPAPAPAPAPAPAPAPAPAPAPAPAP 595
 238 APAPAPAGEAPAPAPAGEAPAPAPAPAPAPAPAPAPAPAPAPAPAPAPAPAPAP 297
 596 TSAPAVAPTPTP-----TPAVAAQEVPA-----SPATGPGPHLSLPTCNPD--- 641
 298 APAPAPAGEAPAPAPAGEAPAPAPAGEAPAPAPAPAPAPAPAPAPAPAPAPAPAP 357
 642 KTD 644

Db 358 RTD 360
 RESULT 4
 NFM_CHICK
 ID NFM_CHICK STANDARD; PRT; 857 AA.
 AC P16053;
 DT 01-APR-1990 (Rel. 14, Created)
 DT 01-APR-1990 (Rel. 14, Last sequence update)
 DT 15-MAR-2004 (Rel. 43, Last annotation update)
 DE Neurofilament triplet M protein (160 kDa neurofilament protein)
 DE (Neurofilament medium polypeptide) (NF-M).
 GN NFM.
 OS Gallus Gallus (Chicken).
 CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 CC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
 CC Gallus.
 OX NCBI_TaxID=9031;
 RN [1]
 RF SEQUENCE FROM N.A.
 RA MEDLINE=90174973; PubMed=2106668;
 RA Zopf D., Dineva B., Betz H., Gundelfinger E.D.;
 RT "Isolation of the chicken middle-molecular weight neurofilament
 RT (NF-M) gene and characterization of its promoter.";
 RL Nucleic Acids Res. 18:521-529(1990).
 RN [2]
 RF SEQUENCE OF 259-857 FROM N.A.
 RA MEDLINE=88112814; PubMed=3123320;
 RA Zopf D., Hermans-Borgmeyer I., Gundelfinger E.D.; Betz H.;
 RT "Identification of gene products expressed in the developing chick
 RT visual system: characterization of a middle-molecular-weight
 RT neurofilament cDNA.";
 RL Genes Dev. 1:699-708(1987).
 CC -!- FUNCTION: Neurofilaments usually contain three intermediate
 CC filament proteins: L, M, and H which are involved in the
 CC maintenance of neuronal caliber.
 CC -!- PTM: There are a number of repeats of the tripeptide K-S-P. NFM is
 CC phosphorylated on a number of the serines in this motif. It is
 CC thought that phosphorylation of NFM results in the formation of
 CC intermediate cross bridges that are important in the maintenance
 CC of axonal caliber.
 CC -!- PTM: Phosphorylation seems to play a major role in the functioning
 CC of the larger neurofilament polypeptides (NF-M and NF-H), the
 CC levels of phosphorylation being altered developmentally and
 CC coincident with a change in the neurofilament function.
 CC -!- SIMILARITY: Belongs to the intermediate filament family.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

EMBL; X17102; CAA34958.1; -;
 EMBL; X05558; CAA29073.1; -;
 PIR; S15762; S15762.
 InterPro; IPR006821; Filament_head.
 InterPro; IPR001664; IF.
 InterPro; IPR002957; Keratin_I.
 Pfam; PF00038; Filament; 1.
 Pfam; PF04732; filament head; 1.
 PRINTS; PR01248; TYPEIKERATIN.
 PROSITE; PS00226; IF; 1.
 Intermediate filament; Coiled coil; Neurone; Phosphorylation;
 Glycoprotein.
 INIT_MET 0 0
 DOMAIN 1 98 HEAD.
 DOMAIN 99 406 ROD.
 DOMAIN 407 857 TAIL.
 DOMAIN 99 130 COIL 1A.


```
CONFLICT 651 656 MISSING (IN REF. 3).
CONFLICT 811 811 E -> A (IN REF. 1).
SEQUENCE 1026 AA; 112478 MW; 087986A08D208C17 CRC64;

Query Match
Best Local Similarity 7.3%; Score 246; DB 1; Length 1026;
Matches 150; Conservative 121; Mismatches 299; Indels 154; Gaps 27;

44 EGYCKVARIQKVETIQRVSEAYENLVKSSKREALEKAMRNKLEGEIRMH----- 97
85 EGCWAVATSRSEKEQLOALNDRFAGYIDKVRQLEAHNR-----LEGEAAALRQQAQGRS 140
98 -----FNRLRE-----RLETKQLAKSEYEGSEDYTKTSQLPFAKNKESQREKEKLE 146
141 AMGELYEREVREVRGAVLRGARGQRLQEHLLED-----IAHVQRLLDDARQREAE 196
147 AEL-ATARNSTNEDQRRHIERDOA-----LSNAQAQKVVKLE 181
197 AAARALARFAQEAARAVDLQKQAQALBECGYLRHRHQBEVGLLGGIQTGSGAAQAQMQ 256
182 BELKKQVYVDKVEKQQAQALVOLQA-ACEKRLQLEHLRLTRLERELSLRQORQNCQP 240
257 AE-TRDLAKCDVTSALREIRAQLEGHAVQSTLQSBFWFRVLDRLSEAAKVNNTDAMRSAQ 315
241 TNVSEYNAALMELLREKEERILALEA-----DMTKWEQKYLEENVVRHFDALDAATVAA 295
316 EETETXR-----HQLQARTTELEALKSTKDSLERQSRSELED---RHQADIASYQEAR 364
296 QROTTVISHSPTSVDTALEARIQKSEEBELMANKKCLDME----- 336
365 QQ-----LDABLNTKWNMAAQLR-----EYQDLNLVKNWALDIRIAAYRKLLEGECRIGFGP 417
337 -----GRITKTHAQI-IEKDAKIKVLQQRSRKPSKTBQLSCMRPAKSLMSINAG 386
418 IPELPEGLPKIPSVSTHIVKSEKIKV-EKSEKETVIVE----- 459
387 SGLSHSSITLGPIMEEKRDDKSWKSGILLLGGDYR-----AEVVPSTPSVPSTPL 442
460 ---QTETQVTEVTEEBEKEEERKEE-EGGESEAEAGGEETKSPFAEASPEK 514
443 SAUSKGTGRDCSTQTERGTESKNTAAVAPISVPAPVAAAATAAATATATATTTMTWAAA 502
515 EAKSPVKEAKSPAEKSPKEAKSPAVKSPKPAKSPAKSPKEAKSPKEAKSPKEAKS 574
503 PVAVAA---AAAPAAAPAAAPATAATAAASVPA-AAGQIPAAASVAGASAAVAPSAAAA 558
575 PAEVKSPKPAKSPAEKSPAEKSPKPAKSPVKEAKSPAEKSPAEKSPAEKSPAEKSP 630
559 AVQVAPAAPVAPALVPVPA-----PAAQAASAPAQTAQTSAPAVAPTAPTPT 610
631 EVKSPKPAKSPKPAKSPKPAKSPKPAKSPKPAKSPKPAKSPKPAKSPKPAKSPKPAKSP 690
611 PAVAQA-EVPASP-----ATGPGPHRLSIPSLTCNPDKTDGPV-PHSNTILR-KTPIQILG 663
691 KAKAPKPAKSPVKEAKSPKPAKSPKPAKSPKPAKSPKPAKSPKPAKSPKPAKSPKPAKSP 750
664 QEPD 667
751 KSPE 754
```

```
SUJT. 7
R_HUMAN
P12270;
01-OCT-1989 (Rel. 12, Created)
01-OCT-1996 (Rel. 34, Last sequence update)
15-MAR-2004 (Rel. 43, Last annotation update)
Nucleoprotein TPR.
Homo sapiens (Human).
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
```

```
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=93064711; PubMed=1437155;
RA Mitchell P.J., Cooper C.S.;
RT "The human tpr gene encodes a protein of 2094 amino acids that has
RT extensive coiled-coil regions and an acidic C-terminal domain.";
RL Oncogene 7:2329-2333(1992).
RN [2]
RP REVISIONS, AND CHARACTERIZATION.
RX MEDLINE=95096166; PubMed=7798308;
RA Byrd D.A., Sweet D.J., Pante N., Konstantinov K.N., Guan T.,
RA Sapshire A.C.S., Mitchell P.J., Cooper C.S., Aebi U., Gerace L.;
RT "Tpr, a large coiled coil protein whose amino terminus is involved in
RT activation of oncogenic kinases, is localized to the cytoplasmic
RT surface of the nuclear pore complex.";
RL J. Cell Biol. 127:1515-1526(1994).
RN [3]
RP SEQUENCE OF 1-342 FROM N.A.
RX MEDLINE=88262257; PubMed=3387099;
RA King H.W.S., Tempest P.R., Merrifield K.R., Rance A.J.;
RT "tpr homologues activate met and raf.";
RL Oncogene 2:617-619(1988).
CC -1- FUNCTION: Component of the cytoplasmic fibrils of the nuclear pore
CC complex implicated in nuclear protein import. Its amino terminus
CC is involved in activation of oncogenic kinases.
CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC SURFACE OF THE NUCLEAR PORE
CC COMPLEX. THE ASSEMBLY OF THE NPC IS A STEPWISE PROCESS IN WHICH
CC TPR-CONTAINING PERIPHERAL STRUCTURES ASSEMBLE AFTER OTHER
CC COMPONENTS, INCLUDING P62.
CC -1- TISSUE SPECIFICITY: Highest in testis, lung, thymus, spleen and
CC brain, lower levels in heart, liver and kidney.
CC -1- DISEASE: Involved in tumorigenic rearrangements with the MET, TRK
CC or RAP genes.
CC -1- DATABASE: NAME=Atlas Genet. Cytogenet. Oncol. Haematol.;
CC WWW="http://www.infobiogen.fr/services/chronocancer/Genes/TPRID282.html".
CC -----
CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; X66397; CAA47021.1; -
CC GO; GO:0005843; C:nuclear pore; TAS.
CC EMBL; Y00672; CAA68681.1; -
CC Genew; HGNC:12017; TPR.
CC -----
CC MIM; 189340; -
CC GO; GO:0005737; C:cytoplasm; TAS
CC GO; GO:0005843; C:nuclear pore; TAS.
CC GO; GO:0006606; P:protein-nucleus import; TAS.
CC Coiled coil; Proto-oncogene; Chromosomal translocation;
CC Nuclear protein; Transport.
CC -----
CC FT DOMAIN 78 360 COILED COIL (POTENTIAL).
CC FT DOMAIN 422 571 COILED COIL (POTENTIAL).
CC FT DOMAIN 575 628 COILED COIL (POTENTIAL).
CC FT DOMAIN 758 805 COILED COIL (POTENTIAL).
CC FT DOMAIN 834 869 COILED COIL (POTENTIAL).
CC FT DOMAIN 934 979 COILED COIL (POTENTIAL).
CC FT DOMAIN 1004 1064 COILED COIL (POTENTIAL).
CC FT DOMAIN 1138 1166 COILED COIL (POTENTIAL).
CC FT DOMAIN 1195 1241 COILED COIL (POTENTIAL).
CC FT DOMAIN 1262 1304 COILED COIL (POTENTIAL).
CC FT DOMAIN 1354 1434 COILED COIL (POTENTIAL).
CC FT DOMAIN 1476 1595 COILED COIL (POTENTIAL).
CC FT DOMAIN 527 530 POLY-SER.
CC FT DOMAIN 1833 1836 POLY-GLU.
CC FT DOMAIN 1957 1964 POLY-ASP.
CC FT DOMAIN 2295 2298 POLY-SER.
CC FT DOMAIN 2349 AA; 265600 MW; AFDD6885CEDCA9EF CRC64;
CC SEQUENCE
CC -----
Query Match 7.3%; Score 245.5; DB 1; Length 2349;
```



```

231 IQQKQGNQCTNVSEYNAAL----- 251
: : : : :
394 VDDITANMEQTVGKANAELWGYEDHNLNETKLDLMDLNTQTKLOBENGEP 453
: : : : :
252 MELLREKEERILALEADMTKEQ-----KYLEENVGRHFALDAATAVAARDTTVISHS 305
: : : : :
454 VRQLEEKESLISQSRGKTSFTQIIEELRQLEBEETKSNAL-AHALQARHDCDLRSEQ 512
: : : : :
306 P-----NTSYDT-----ALARIQKEEIEILM 327
: : : : :
513 YEEFOEAKELQALSKGNAEVAQWTKYETDAIORTEBEDAKKLLARLQEAEEAIEA 572
: : : : :
328 ANKRCCLDMGRITKLH-AQIIEKDAI-----KVLQQRSRK-RPS 365
: : : : :
573 ANAKCSSLE---KAKHRLQNEQEDMIDLEKANSAAASLDKQKQFGDKIINDWKQKYES 629
: : : : :
366 KTEQLSCMRPAKS-----LMSISNAGSGLLSHSTLTGSPIMEKRDKS 410
: : : : :
630 QAELEASQKEARSLSITELFKLNAYEETLDHLETL-----KRENKN 670
: : : : :

SNAP91.
Rattus norvegicus (Rat).
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
NCBI_TaxID=10116;
[1]
SEQUENCE FROM N.A. (ISOFORMS LONG AND SHORT).
TISSUE=Brain;
MEDLINE=93178442; PubMed=8440257;
Morris S.A., Schroeder S., Plessmann U., Weber K., Ungewickell E.;
"Clathrin assembly protein Ap180: Primary structure, domain
organization and identification of a clathrin binding site.";
EMBO J. 12:667-675 (1993).
-!- FUNCTION: Adaptins are components of the adaptor complexes which
link clathrin to receptors in coated vesicles. Clathrin-associated
protein complexes are believed to interact with the cytoplasmic
tails of membrane proteins, leading to their selection and
concentration. Binding of Ap180 to clathrin triskelia induces
their assembly into 60-70 nm coats.
-!- SUBCELLULAR LOCATION: Component of the coat surrounding the
cytoplasmic face of coated vesicles in the plasma membrane.
-!- ALTERNATIVE PRODUCTS:
Event=Alternative splicing; Named isoforms=2;
Name=Long;
IsoId=Q05140-1; Sequence=Displayed;
Name=Short;
IsoId=Q05140-2; Sequence=VSP_000173;
-!- DOMAIN: Possesses a three domain structure: the N-terminal 300
residues harbor a clathrin binding site, an acidic middle domain
450 residues, interrupted by an Ala-rich segment, and the C-
terminal domain (166 residues).
-!- PTM: Phosphorylated (By similarity).
-!- SIMILARITY: Contains 1 epsin N-terminal homology (ENTH) domain.

This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@sib-sib.ch).
EMBL; X68877; CAA48748.1; -.

```

```

DR EMBL; X68878; CAA48749.1; -.
DR PIR; S36326; S36326.
DR PIR; S36327; S36327.
DR InterPro; IPR001026; ENTH.
DR InterPro; IPR008943; PI_bind_N.
DR Pfam; PF01417; ENTH; 1.
DR SMART; SM00273; ENTH; 1.
DR PROSITE; PS00942; ENTH; 1.
KW Coated pits; Alternative splicing; Phosphorylation.
FT DOMAIN 14 145 ENTH.
FT DOMAIN 410 413 POLY-THR.
FT DOMAIN 535 539 POLY-ALA.
FT DOMAIN 547 550 POLY-ALA.
FT DOMAIN 678 683 POLY-SER.
FT DOMAIN 723 729 POLY-SER.
FT VARSPIC 614 632 Missing (in isoform Short).
FT FTIDEVSP_000173.
SQ SEQUENCE 915 AA; 93518 MW; 32ECLB38EC5DF8C0 CRC64;

Query Match 7.2%; Score 242; DB 1; Length 915;
Best Local Similarity 21.0%; Pred. No. 0.0074;
Matches 145; Conservative 98; Mismatches 282; Indels 166; Gaps 26;

QY 28 MVEILSDENLNLEQELGEGYKVKARLOKVEIQRVSEAYENLVKSSSKREALEKAMRN- 86
DB 45 LIQATNETNVI PQADTLPERATNSWV--VVKALAVTHLHVMHGNERFIQYLASRNT 102
QY 87 --KLGEGIRWHDNDRLERLETANKQLAEKEY-----EGSEDTRKTIISQL 131
DB 103 LFNLSNFDLSGSHGDMSTFIRYSRYLNEKAPSYRQMAFDFARVKKGADGVNRN--- 159
QY 132 FAKNKESQREKEKLEALATARSTNEDQRHRIIRDQALSNA---QAKVYKLEELKKQ 188
DB 160 -----VPEKLLKSMPTIQG-QIDALLEPDPVHNPDLTNGVINAFAFLLFKDLIKLEA 209
QY 189 VYVDKYEKMOQALVOL-QAACEKGEQLEHRLRLRELESLESLRIQORQNCQPTNVSEYN 247
DB 210 CYNDGVNLLLEKFEKKGCKDALSIYKFLTRMTRVSEPLKVADEVG-IDKGDIPDLT 268
QY 248 AA--ALMELLREKEERILALEADMTKEQKYLEENVGRHFALDAATAVAARDTTVISHS 305
DB 269 QAPSSLMETL---EQHLNLTLEG-----KXPGNNEGSGAPSPSKSPATTVT-S 313
QY 306 PNTSYDTALEARIQKEEIEILMANKCLDMGRITKLHQAIIKEDAMIKVLQQRSRKPS 365
DB 314 PNST-----PAKTIIDTSPFVDIFATASAAAP- 339
QY 366 KTEQLSCMRPAKSLMSTISNAGSGLLSHSTLTGSPIMEKRDKSKWSGLIGLLGQDYRA 425
DB 340 ----VSSAKSSDLLLDLPDFSCARAGAA----APVPTGGATAM----GDLGSDSLA 387
QY 426 EYVPTSPVPVPTPLLSAHSKNTGSDCSTQTERGTESNKTAAPISVPAPVAAATAA 485
DB 388 A-LSSVSEAPISDPFAPESPSP-----TTTTT-----PASASASATTA 425
QY 486 AITATAATITTTWVAAPVAVAAAAAATAATAATAATAATAATAATAATAATAATAATA 545
DB 426 ---VTAATTEVDLFGDPAPASPGEAAPASGATAPATPAPVAAALD-ACSENDPPAPSEG 481
QY 546 SAA-----AVAPSAASAAAAV-----QVAPAPAPVPPAPALVEVPAPAAQASAPAQTO 593
DB 482 SAEAAPDLDFAMKPPETSAPVVTPTASTAPPVPATAPSPAPTAVATAATTAATTTAAATAAT 541
QY 594 APTGAPAVATPTAPTP-----TPAVAAQAYP-----ASPATGKGP 628
DB 542 TATTSAATAATAAPPALDIFGDLFDASPEVAASKEDVAPSIDLFGTDAFSPFPGASP 601
QY 629 HRLISPLSTCNPKPTDGFVPHSNLTLEKPTPI 659
DB 602 ----VPSSSLTADLLSGSGFHCAEDDRHVPFL 628

```

RESULT 10

Dautigny A., Pham-Dinh D., Roussel C., Felix J.M., Nussbaum J.L., Jolles P.;
"The large neurofilament subunit (NF-H) of the rat: cDNA cloning and in situ detection";
Biochem. Biophys. Res. Commun. 154:1099-1106(1988).

[3]

SEQUENCE OF 1-89 AND 243-313 FROM N.A.

MEDLINE=87080760; PubMed=2878828;

Robinson P.A., Wion D., Anderton B.H.;

"Isolation of a cDNA for the rat heavy neurofilament polypeptide (NF-H)";

FEBS Lett. 209:203-205(1986).

[4]

SEQUENCE OF 318-831 FROM N.A.

MEDLINE=89184647; PubMed=2928342;

Lieberburg I., Spinner N., Snyder S., Anderson J., Goldgaber D.,

Smulowitz M., Carroll Z., Emanuel B.S., Breitner J., Rubin L.;

"Cloning of a cDNA encoding the rat high molecular weight

neurofilament peptide (NF-H): developmental and tissue expression in

the rat, and mapping of its human homologue to chromosomes 1 and

22.";

Proc. Natl. Acad. Sci. U.S.A. 86:2463-2467(1989).

-!- FUNCTION: Neurofilaments usually contain three intermediate

filament proteins: L, M, and H which are involved in the

maintenance of neuronal caliber. NF-H has an important function in

mature axons that is not subverted by the two smaller NF proteins.

-!- PTM: There are a number of repeats of the tripeptide K-S-P. NFH is

phosphorylated on a number of the serines in this motif. It is

thought that phosphorylation of NFH results in the formation of

interfilament cross bridges that are important in the maintenance

of axonal caliber.

-!- PTM: Phosphorylation seems to play a major role in the functioning

of the larger neurofilament polypeptides (NF-M and NF-H), the

levels of phosphorylation being altered developmentally and

coincident with a change in the neurofilament function.

-!- SIMILARITY: Belongs to the intermediate filament family.

-!- CAUTION: Ref.1 sequence differs from that shown due to a

frameshift in position 783.

This SWISS-PROT entry is copyright. It is produced through a collaboration

between the Swiss Institute of Bioinformatics and the EMBL outstation

the European Bioinformatics Institute. There are no restrictions on its

use by non-profit institutions as long as its content is in no way

modified and this statement is not removed. Usage by and for commercial

entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

or send an email to license@isb-sib.ch).

EMBL; M37227; AAA41693.1; ALT_FRAME.

EMBL; X13804; CAA32038.1; ALT_FRAME.

EMBL; M21964; AAA41695.1; -.

EMBL; J04517; AAA41692.1; -.

PIR; S02003; S02003.

InterPro; IPR001664; IF.

Pfam; PF00038; Filament; 1.

PROSITE; PS00226; IF; 1.

InterMediate filament; Coiled coil; Neurone; Phosphorylation; Repeat.

NON_TER 1 1

DOMAIN 276 641

CONFLICT 164 164

CONFLICT 185 185

CONFLICT 193 193

CONFLICT 199 199

CONFLICT 346 346

CONFLICT 373 373

CONFLICT 482 482

CONFLICT 485 485

CONFLICT 570 571

CONFLICT 591 591

CONFLICT 727 727

CONFLICT 757 759

CONFLICT 769 769

CONFLICT 775 775

CONFLICT 831 AA; 89486 MW; 1B0973C3F13EF768 CRC64;

SEQUENCE

Query Match 7.2%; Score 241.5; DB 1; Length 831;
Best Local Similarity 22.4%; Pred. No. 0.007;
Matches 131; Conservative 87; Mismatches 199; Indels 169; Gaps 26;

```
QY 168 QALSNQAQKVKLEELKKQVYVYKVEKQQAALVQLQA-ACREKQLEHLRLRLEREL 226
DB 19 QAQAQAAR-----DALK-----CDVTSALRRLRAQLEHTVQSTLOSEFWFVRLDRLS 68
QY 227 ESLRIQOQGNQCPNTNVEVNAALMELLREKEERILALEA-----DMTKWEQKYLEENV 281
DB 69 EAAKVNTDAMESAQSEITEYK-----RQLQARTTELEALKSTKESLERQRSELED-- 118
QY 282 MEHFALDAAATVAAQORTVTYISHSPNTSYDTALEARLOKSEEEILMANKCLDWE----- 336
DB 119 -RH-----QVDMASYQD--AIQQLDNELRNTKMAAQLREYQDLNFKMALDLUEIAAYR 170
QY 337 -----GRITLHAQI--IENDAMIKVLQORSRKE-----PSKTE 368
DB 171 KLLBGEBCRIGFGPIPTSLTGLPKIPSMSTHIKVXSEKIKV--EKSEKETVIVERQTE 229
QY 369 QLSMRPAKSLMSISNAGSGLLSHSSTLTGSPINER--KRDDKSWKG--SLGILLGGDYR 424
DB 230 BIQ-----VTEVTEEDKEAQGEEREBEAEEGEEA 260
QY 425 AEYVPTSPVPVPTPLLSAHSKGTGRDCSTQTERGTESN-KTAAVAPISVPAPVAAAAT 483
DB 261 ATTSPPAEEAASP-----EKETKSPVKEEKSPAEAKSPAEAKSP 300
QY 484 AAAITATATATITTTWAAAPVAV-----AAAAAATAAAT 530
DB 301 AEA--KSPAEVKSAPVAKSPAEVKSAPAEVKSAPAEVKSAPAEVKSAPAEVKSAPAEVKS 358
QY 531 SPAAAGQIPA-----AASVASAAVAPSAASAAAQVQVAPA--APAPVPAPALV----- 576
DB 359 SPAEA-KSPAEEVKSAPAEVKSAPAEVKSAPAEVKSAPAEVKSAPAEVKSAPAEVKSAPAEV 417
QY 577 -----PVFPAPAAQAASAPAQAPTS-----APAVAPTPAPTETPAPVQAQVPA 620
DB 418 KSPAEEVKSAPVTVKSAPAEVKSAPAEVKSAPAEVKSAPAEVKSAPAEVKSAPAEVKSAPAEV 477
QY 621 SPATGPGPHRLSLTSLTCNPKDTPGVFHSNTLERTPIQLQGE 665
DB 478 KSPAGAKPPAEAKSPAEAKSPAEAKSPAEAKSPAEAKSPAEAKSPAEAKSPAEAKSPAEAKSP 523
```

RESULT 12

```
CPN_DROME STANDARD; PRT; 865 AA.
AC Q02910;
DT 01-OCT-1993 (Rel. 27, Created)
DT 01-OCT-1993 (Rel. 27, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Calphorin.
GN CPN OR CAP.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Sphingidae; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]_TaxID=7227;
RP SEQUENCE FROM N.A.
RC STRAIN=Canton-S;
RA MEDLINE=93165729; PubMed=8094559;
RX Martin J.H., Benzer S., Rudnicka M., Miller C.A.;
RT "Calphorin: a Drosophila photoreceptor cell calcium-binding protein.";
RL Proc. Natl. Acad. Sci. U.S.A. 90:1531-1535(1993).
RN [2];
RP SEQUENCE FROM N.A.
RC STRAIN=Canton-S;
RA MEDLINE=93165730; PubMed=8434015;
RX Ballingier D.G., Xue N., Harshman K.D.;
RT "A Drosophila photoreceptor cell-specific protein, calphorin, binds
```


243 VSEYNAAALAKELLERERILALEAD-----MTWEQKYLEEN--VMRH:FALDAATV 293
RA Sandelin A., Schneider C., Sempie C.A., Setou M., Shimada K.,
RA Sultana R., Takenaka Y., Taylor M.S., Teasdale R.D., Tomita M.,
324 ETEHEEPQWELLAKHKTITTSREADEGTIDIEKQFLDBYTTKENQKVEESQADARGEN 383
RA Verardo R., Wagner L., Wahlestedt C., Wang Y., Watanabe Y., Wells C.,
RA Wilming L.G., Wyshaw-Boris A., Yanagisawa M., Yang I., Yang L.,
294 AQRDITVISHSNTSYDTALEARIQKEEELIMANKRCLDMGRKTLHAQ----- 345
RA Yuan Z., Zavalan M., Zhu Y., Zimmer A., Carninci P., Hayateu N.,
RA Hirozane-Kishikawa T., Konno H., Nakamura M., Sakazume N., Sato K.,
RA Shiraki T., Waki K., Kawai J., Aizawa K., Arakawa T., Fukuda S.,
RA Hara A., Hashizume W., Imotani K., Ishii Y., Iton M., Kagawa I.,
RA Miyazaki A., Sakai K., Sasaki D., Shibata K., Shinagawa A.,
RA Yaenishi A., Yoshino M., Waterston R., Lander E.S., Rogers J.,
346 -----IIEKAMIKVLOQRSKPEKTEQLSCMRPA----- 376
RA Birney E., Hayashizaki Y.,
348 NPFQMASGVGNKSEEPKQKHKEEBEPEQLQELPALPMPFVDPSSNPPFRKSNLS 497
RT "Analysis of the mouse transcriptome based on functional annotation of
RT Nature 420:563-573(2002)."
377 -----KSLMSISNAGSLLSHSSTLTGSPIMEEKR 406
CC phosphatidylinositol 3-phosphate and participates in endosomal
CC trafficking (By similarity).
CC RAB22A that have been activated by GTP-binding (By similarity).
CC RAB22A SUBUNIT: Homodimer. Binds STX6. Binds RAB5A, RAB5B, RAB5C and
CC associated with early endosomes (By similarity).
CC associated with early endosomes (By similarity).
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC at the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AK046231; BAC32647.1; -
KW MG1; 2442192; Real.
KW Coiled coil. 1 789 COILED COIL (POTENTIAL).
FT DOMAIN 180 550 GLN/GLU/LYS-RICH.
FT NON TER 790 790
FT SEQUENCE 790 AA; 90539 MW; A167AB02CC2CAYBF CRC64;
SQ
Query Match 7.1%; Score 239; DB 1; Length 790;
Best Local Similarity 20.3%; Pred. No. 0.0083;
Matches 127; Conservative 96; Mismatches 187; Indels 216; Gaps 22;
QY 15 PADPFAIVSRAQM---VEILSDENRHLRQEL----- 44
DB 124 PQEVAVTVQETQKLKGSINBLTKQNLTKLQKLDLYTHLEEKNEESASRKTQASL 183
QY 45 -----GCYKVARLQKVETETQVSEAYENLVKSSSKREALEKAMRNKLEGERMHDN 99
DB 184 HORLDCQQLQARLTASESSLQRAQ-----ELSEKAEAAQ-----KLREELREVESTR 232
QY 100 RDLRERLETANKLAEEYEGSEDTTKTISQIFAKNKVESOREKLEAELEATARNEDQ 159
DB 233 QHLKVEVKQLQQQREKXEQHGLQ--LQGEVSQHLCKLLETERQLGEAHGRLEKQRLSSEK 291
QY 160 RREHTEIRDOALSNAQKVKVLEELKKK-----QVYVDVKVKKQQAALVOLQAACEK 210
DB 292 ---LMEKEQVQVADLQKLRLLEEQKEKVTNSTELOHLEKSKQHQEQALQASATAKL 348
QY 211 RE---QLEHLR-----TLRELESL---RIQROGNCQPTN 242
DB 349 REAQNDLEQVLRQIGDKQDKQNTQNDLEALLQKGSVSLLEKXEDLYAKIQAGEGETAVLN 408
QY 243 -VSEYNAALMEL-----LREKERILALEADMT 270
DB 409 QLOEKHQAQQLTQLEKLNQSESHQABENLHDQVQKAKHLRAAQRVLSLETSSVS 458
QY 271 -----KWEQKYLEENVMRH:FALDAATVAAQR----- 297
DB 469 ELSQINESKEKVSQIDIQIKAKTELLLSAEAKAAQADLQNHLDTAQHALQKQOELN 528
QY 298 -----DTTVISHSNTSYDTALEARIQKEEELIMANKRCLDMGRKTLHAQIE--- 348

SUIT 14

A1 MOUSE
Q8BL56:
10-OCT-2003 (Rel. 42, Created)
10-OCT-2003 (Rel. 42, Last sequence update)
10-OCT-2003 (Rel. 42, Last annotation update)
Early endosome antigen 1 (fragment).
EEL1.
Mus musculus (Mouse).
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
NCBI_TaxID=10090;
[1]
SEQUENCE FROM N.A.
STRAIN=C57BL/6J; TISSUE=Brain;
MEDLINE=22354683; PubMed=12466851;
Okazaki Y., Furuno M., Kasukawa T., Adachi J.J., Bono H., Kondo S.,
Nikaido I., Oatso N., Saito R., Suzuki H., Yananaka I., Kiyosawa H.,
Yagi K., Tomaru Y., Hasegawa Y., Nogami A., Schenbach C., Gojohori T.,
Baldarelli R., Hill D.P., Bult C., Hume D.A., Quackenbush J.,
Schriml L.M., Kanapin A., Matsuda H., Batalov S., Beisel K.W.,
Blake J.A., Bradt D., Brusic V., Chothia C., Corbani L.E., Cousins S.,
Dalla E., Dragani T.A., Fletcher C.F., Forrest A., Frazer K.S.,
Gaasterland T., Gariboldi M., Giesi C., Godzik A., Gough J.,
Grimmond S., Gustincich S., Hachawa N., Jackson I.J., Jarvis E.D.,
Kanai A., Kawaji H., Kawasawa Y., Kedzierski R.M., King B.B.,
Konagaya A., Kurochkin I.V., Lee Y., Lenhard B., Lyons P.A.,
Maglott D.R., Maltais L., Marchionni L., McKenzie L., Miki H.,
Nagashima T., Numata K., Okido T., Pavan W.J., Pertea G., Pesole G.,
Petrovsky N., Pillai R., Pontius J.U., Qi D., Ramachandran S.,

```

529 KVSQVQLDGLTAKFQKQEHCIQLESKDKHKKHLSLEQKVEDLEGHIKKLEADALEYKA 588
349 -KDAIKVLYOR-----SRKPSKTEQLSCMRPAKSLMSISNAGSLGHS 393
589 SKEQLQSQQQQLSTDLRLNALSRLQEQEVVSCYK-----LDLQNKSEILENIX 643
394 STLTGSPIMEKRDCKWKGSLGILLGDYRABYVFPSTPSPPLLSAHSKT-----448
644 QTLT-----KKEEN-----VVLKQEFK-----LSQSKTQHKEL 674
449 GSRDCSTOTE-RGTESNKTAAVAPIS 473
675 GDRMQAAVETLTAVKQDKALLAELS 700

35015
180 MOUSE
A180 MOUSE STANDARD; PRT; 901 AA.
Q61548; O61547;
01-NOV-1997 (Rel. 35, Created)
01-NOV-1997 (Rel. 35, Last sequence update)
15-MAR-2004 (Rel. 43, Last annotation update)
Clathrin coat assembly protein APL80 (Clathrin coat associated protein
APL80) (91 kDa synaptosomal-associated protein) (Phosphoprotein P1-
20).
SNAP91.
Mus musculus (Mouse).
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
NCBI_TaxId=10090;
[1]
SEQUENCE FROM N.A. (ISOFORMS LONG AND SHORT).
MEDLINE=92300439; PubMed=1607933;
Zhou S., Sousa R., Tannery N.H., Lafer E.M.;
"Characterization of a novel synapse-specific protein. II. CDNA
cloning and sequence analysis of the 21-20 protein.";
J. Neurosci. 12:2144-2155(1992).
-!- FUNCTION: Adaptins are components of the adaptor complexes which
link clathrin to receptors in coated vesicles. Clathrin-associated
protein complexes are believed to interact with the cytoplasmic
tails of membrane proteins, leading to their selection and
concentration. Binding of APL80 to clathrin triskelia induces
their assembly into 60-70 nm coats.
-!- SUBCELLULAR LOCATION: Component of the coat surrounding the
cytoplasmic face of coated vesicles in the plasma membrane.
-!- ALTERNATIVE PRODUCTS:
Event=Alternative splicing; Named isoforms=2;
Name=Long;
IsoId=O61548-1; Sequence=Displayed;
Name=Short;
IsoId=O61548-2; Sequence=VSP_000172;
-!- TISSUE SPECIFICITY: Brain. Associated with the synapses.
-!- DEVELOPMENTAL STAGE: Developmentally regulated in a pattern
coincident with active synaptogenesis and synaptic maturation.
-!- DOMAIN: Possesses a three domain structure: the N-terminal 300
residues harbor a clathrin binding site, an acidic middle domain
450 residues, interrupted by an Ala-rich segment, and the C-
terminal domain (166 residues).
-!- PTM: Phosphorylated.
-!- SIMILARITY: Contains 1 epsin N-terminal homology (ENTH) domain.

This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/
or send an email to license@isb-sib.ch).
-----
EMBL; M83985; AAA37587.1; ..
EMBL; M83985; AAA37586.1; ..
PIR; A44825; A44825.
MGD; MGI:109132; Snap91.

```

```

DR InterPro; IPR001026; ENTH.
DR InterPro; IPR008943; PI_bind_N.
DR Pfam; PF01417; ENTH; 1.
DR SMART; SM00273; ENTH; 1.
DR PROSITE; PS0942; ENTH; 1.
KW Coated pits; Alternative splicing; Phosphorylation.
FT DOMAIN 14 145
FT DOMAIN 14 145
FT DOMAIN 410 413
FT DOMAIN 535 539
FT DOMAIN 547 550
FT DOMAIN 559 564
FT DOMAIN 659 664
FT DOMAIN 704 710
FT VARSPLIC 715 719
FT Missing (in isoform Short).
FT /FTID=VSP_000172.
SQ SEQUENCE 901 AA; 91851 MW; 24A98FBACEBDB831 CRC64;

Query Match 7.0%; Score 234; DB 1; Length 901;
Best Local Similarity 20.8%; Pred. No. 0.015;
Matches 152; Conservative 105; Mismatches 281; Indels 192; Gaps 30;

QY 28 MVEILSDENRNLRQELGCGYKVARLQKVFETIORVSEAYENLVKSSSKREALEKAVRN- 86
DB 45 LIQATNETNVTIPQWADTLFERATNSSWV--VFKALVTTTHLMVHGNERFIQYLSRNT 102
QY 87 --KLEGEIRMHEDFNRLRERLETANKOLAEKY-----EGSEDTKRTIISQL 131
DB 103 LFNLSNFTLKGSGHYDMSFTIRYSRYLAEKAPSYQMAFDPAVKKGAGVWRTM--- 159
QY 132 FAKNKESREKEKLEALATARNSTNEQRRHIEIRDOALSNA---QAKVVKLEELKKQ 188
DB 160 -----VPEKLLKSMPIQG-QIDALLEFDVHPNELNGVINAAPMLLFKDLIKLFA 209
QY 189 VYVDVKEMQOALVOL-QAAEKEQLEHLRLTRLESELESRLIQORQGNQCPTNVSEYN 247
DB 210 CYNDGVINLLEKFFPMKGGQCKDALEYKFLTMTVSEFLKVAEQVG-IDKGDIPDLT 268
QY 248 AA--ALMELLKEKERILALRADTKWEQKYLENVNRHFDALDAATVAQRODTTVISHS 305
DB 269 QAPSSIMETL---EQHLNLTLEG-----KKPGNEGSGAPSPLSKSSPAITVT-S 313
QY 306 PNTSYDTALEARIQKEEBEILMANKRCLDMEGRIKTLHAQIIEKDAIKVKLQORSRKEPS 365
DB 314 PNST-----PACTIDTSPVDIFATASAAAP- 339
QY 366 KTEQLSCHRPKAKSLMSISNAGSLSHSSTLTGSPIMEEKRDDKSKWGLGILLGGDYRA 425
DB 340 -----VSSAKPSDLDLQPDFSGAAGAA-----APVVPSPGGATW---GDLLGSDSLA 387
QY 426 EYVPTSPVPSTPLLSAHSKTSGRDCSTQTERGTESNKTAAVAPISVPAVAAATAA 485
DB 388 A-LSSVPCAPISDPF-----APEP-SPPTTTTEPASAS 419
QY 486 ALTATATATITTMVAAPVAVAAA--AAPAAA-AAPSPATAATAAAVSPAAAGQIPAAA 542
DB 420 ASTTTAVTAVTTEVDLFGDAPFAASPGEPAPASGATAPATPAPVPAALD-ACSGNDPFPAP 478
QY 543 SVASAA-----AVAPSAAMAAV-----QVAPAPAPVPAPALVPVAPAPAAQASADA 590
DB 479 SEGSAAPAPDLDFAMKEPPEPSAVTPTASTAPPVATAPSPAPTAAVATAATTAATAA 538
QY 591 QTQAPTSAPAVAPTPTP-----TPVAQAQEV-ASPATG 625
DB 539 ATTTATTTAAATTAAPPALDIFGDLFSAPVAAAPKPDAAAPSIDLFGTDAFSSPPRG 598
QY 626 PGP-----HRLSI-----PSLTCPNDKTD-----QPVFHSNTLE-RKTPIQI 661
DB 599 ASPVPESSLTADLLSVDAFAAPSPASTASPAKAESSGVLDLFGDASGSGASSTQAPAPQAV 658
QY 662 LQEPDAEW 671
DB 659 SSSSASADLL 668

```

arch completed: June 16, 2004, 19:17:44
b time : 20.5037 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

1 protein - protein search, using sw model

in on: June 16, 2004, 19:08:47 ; Search time 62.7139 Seconds
(without alignments)
3395.972 Million cell updates/sec

tle: US-09-332-063-2
rfect score: 3347
quence: 1 MPRAPSSASQPVADPFA.....KTPIQLGQEPDAEMVEYLI 675

oring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

arched: 1017041 seqs, 315518202 residues

tal number of hits satisfying chosen parameters: 1017041

imum DB seq length: 0

imum DB seq length: 2000000000

st-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

atabase :

- SPTREMBL_25.*
- 1: sp_archaea.*
 - 2: sp_bacteria.*
 - 3: sp_fungi.*
 - 4: sp_human.*
 - 5: sp_invertebrate.*
 - 6: sp_mammal.*
 - 7: sp_mhc.*
 - 8: sp_organelle.*
 - 9: sp_phage.*
 - 10: sp_plant.*
 - 11: sp_rodent.*
 - 12: sp_virus.*
 - 13: sp_vertebrate.*
 - 14: sp_unclassified.*
 - 15: sp_rvirus.*
 - 16: sp_bacteriap.*
 - 17: sp_archaeap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

result No.	Score	Match	Length	DB	ID	Description
1	3347	100.0	675	4	Q9HD27	Q9hd27 homo sapien
2	2683	80.2	891	11	Q8VHC2	Q8vhg2 mus musculu
3	1448	43.3	956	4	Q8IY63	Q8ly63 homo sapien
4	1400	41.8	882	11	Q9D4H4	Q9d4h4 mus musculu
5	1338	40.0	498	4	Q96CM5	Q96cm5 homo sapien
6	1090.5	32.6	772	11	Q8K371	Q8k371 mus musculu
7	1087.5	32.5	463	11	Q8BS08	Q8bs08 mus musculu
8	1075.5	32.1	463	11	Q8BP84	Q8bp84 mus musculu
9	1069.5	32.0	586	11	Q9QU50	Q9qu50 mus musculu
10	1066.5	31.9	463	11	Q7TPE4	Q7tpe4 mus musculu
11	1045	31.2	569	4	Q9UKB4	Q9ukb4 homo sapien
12	1040.5	31.1	466	4	Q96F99	Q96f99 homo sapien
13	1040.5	31.1	859	4	Q9Y2J4	Q9y2j4 homo sapien
14	867.5	25.9	402	4	Q8TER8	Q8ter8 homo sapien
15	738.5	22.1	326	4	Q8NDN0	Q8ndn0 homo sapien
16	549.5	16.4	612	4	Q8TEN8	Q8ten8 homo sapien

17	427.5	12.8	557	4	Q8WXD1	Q8wxd1 homo sapien
18	335	10.0	203	11	Q9QZP5	Q9qzp5 rattus norv
19	328.5	9.8	182	4	Q9H879	Q9h879 homo sapien
20	325	9.7	789	16	Q8XX15	Q8xx15 talstonia s
21	310.5	9.3	3326	12	Q7T591	Q7t591 simian herp
22	300	9.0	3288	12	Q7T5D9	Q7t5d9 simian herp
23	296.5	8.9	1013	5	Q9VYT7	Q9vyt7 drosophila
24	294.5	8.8	1346	5	Q9U183	Q9u183 leishmania
25	292	8.7	507	13	Q13028	Q13028 boreogadus
26	292	8.7	1046	11	Q8OTQ3	Q8otq3 mus musculu
27	286.5	8.6	2346	5	Q01385	Q01385 drosophila
28	286	8.5	1069	5	Q8EBG1	Q8ebg1 drosophila
29	285.5	8.5	1072	11	Q35482	Q35482 rattus norv
30	284.5	8.5	2346	5	Q9V624	Q9v624 drosophila
31	282	8.4	1354	11	Q9EPW8	Q9epw8 mus musculu
32	280	8.4	3534	12	Q39266	Q39266 equine herp
33	279	8.3	801	5	Q23635	Q23635 caenorhabdi
34	278	8.3	600	5	Q21036	Q21036 caenorhabdi
35	275	8.2	2701	4	Q9Y520	Q9y520 homo sapien
36	274.5	8.2	1299	4	Q9NSM8	Q9nsm8 homo sapien
37	274	8.2	426	2	Q9LAV5	Q9lays streptococ
38	266.5	8.0	875	2	Q9F280	Q9f2b0 thioacillu
39	266.5	8.0	997	5	Q9W2J2	Q9w2j2 drosophila
40	266	7.9	668	5	Q8MXU8	Q8mxu8 caenorhabdi
41	265	7.9	675	13	Q3W608	Q3w6j8 dissostichu
42	265	7.9	1020	5	Q86PC3	Q86pc3 drosophila
43	258	7.7	1174	4	Q94854	Q94854 homo sapien
44	257.5	7.7	2037	13	Q9PVZ2	Q9pvz2 xenopus lae
45	257	7.7	1480	11	Q80TM9	Q80tm9 mus musculu

ALIGNMENTS

RESULT 1

Q9HD27	PRELIMINARY;	PRT;	675 AA.
ID	Q9HD27	Q9UPT1;	
AC	Q9HD27; Q9UPT1;		
DT	01-MAR-2001 (Tremblrel. 16, Created)		
DT	01-MAR-2001 (Tremblrel. 16, Last sequence update)		
DT	01-OCT-2003 (Tremblrel. 25, Last annotation update)		
DE	Angiomotin (Hypothetical protein KIAA1071).		
GN	KIAA1071.		
OS	Homo sapiens (Human).		
OC	Zukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
OX	NCBI_TaxID=9606;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RX	MEDLINE=21157403; PubMed=11257124;		
RA	Troyanovsky B., Levchenko T., Mansson G., Matvienko O., Holmgren L.;		
RT	"Angiomotin. An angiotensin binding protein that regulates endothelial cell migration and tube formation."		
RL	J. Cell Biol. 152:1247-1254(2001).		
RN	[2]		
RP	SEQUENCE FROM N.A.		
RC	TISSUE=Brain;		
RX	MEDLINE=99397452; PubMed=10470851;		
RA	Kikuno R., Nagase T., Ishikawa K., Hirose M., Miyajima N., Tanaka A., Kotani H., Nomura N., Ohara O.;		
RT	"Prediction of the coding sequences of unidentified human genes. XIV. The complete sequences of 100 new cDNA clones from brain which code for large proteins in vitro."		
RL	DNA Res. 6:197-205(1999).		
DR	ENBL; AF286598; BAG01851.1; -		
DR	ENBL; AB028994; EAA83023.3; -		
KW	Hypothetical protein.		
SQ	SEQUENCE 675 AA; 72540 MW; EBC28B74427AD481 CRC64;		

Query Match 100.0%; Score 3347; DB 4; Length 675;
Best Local Similarity 100.0%; Pred. No. 5.9e-138;
Matches 675; Conservative 0; Mismatches 0; Indels 0; Gaps 0;


```

1 MPRAOPSSASQVPADPFAIVSRAQQWVEILSDENRLROELGCGYKVARLQKVTEI 60
1 MPRAOPSSASQVPADPFAIVSRAQQWVEILSDENRLROELGCGYKVARLQKVTEI 60
61 QVSVAYENLVKSSSKREALSKAMENKLEGEIRRMHDFNRDLRERLETANKQLAEKEYEG 120
61 QVSVAYENLVKSSSKREALSKAMENKLEGEIRRMHDFNRDLRERLETANKQLAEKEYEG 120
121 SDTTRKTTISQLPACKKESQREKELEALATARSNEDORRHIERDOALSNAQ 180
121 SDTTRKTTISQLPACKKESQREKELEALATARSNEDORRHIERDOALSNAQ 180
181 BEELKKKKQVYVDKVKMOQALVOLQAACEKEQLEHRLRTRLELESRLRQFGNCQP 240
181 BEELKKKKQVYVDKVKMOQALVOLQAACEKEQLEHRLRTRLELESRLRQFGNCQP 240
241 TNVSEYNAALMELLREKEERILALEADMTKWEQKYLEENVMRHFALDAATVA 300
241 TNVSEYNAALMELLREKEERILALEADMTKWEQKYLEENVMRHFALDAATVA 300
301 VLSHSPNTSYDTALEARIQKEEBEILMANKRCLDMGRIKTLHAQIIEKDMIKVLQORS 360
301 VLSHSPNTSYDTALEARIQKEEBEILMANKRCLDMGRIKTLHAQIIEKDMIKVLQORS 360
361 RKEPSKTEQLSCWRPAKSLMSISNAGSGLLSHSSTLTGSPIMEEKRDDKSKWGSIGILLG 420
361 RKEPSKTEQLSCWRPAKSLMSISNAGSGLLSHSSTLTGSPIMEEKRDDKSKWGSIGILLG 420
421 GYRAEYVSTPSPVPPTPLLSAHSKTSGRDCSTQTERGTSNKTAAPVAPVAVAA 480
421 GYRAEYVSTPSPVPPTPLLSAHSKTSGRDCSTQTERGTSNKTAAPVAPVAVAA 480
481 AATAAAITATAATITTTWVAAPVAVAAAAAPAAAAAPSPATAATAAAVSPAAQIIPA 540
481 AATAAAITATAATITTTWVAAPVAVAAAAAPAAAAAPSPATAATAAAVSPAAQIIPA 540
541 AASVASAAAASAAAAAVQVAPAPVAPAPVAPVAPVAPVAPVAPVAPVAPVAPVAPV 600
541 AASVASAAAASAAAAAVQVAPAPVAPAPVAPVAPVAPVAPVAPVAPVAPVAPVAPV 600
601 VAPTPTPTTVAQAQVAVASAPATGPGPHRLSIPSLTNCNPKDTCGPFVPHSNTLKRKTIQ 660
601 VAPTPTPTTVAQAQVAVASAPATGPGPHRLSIPSLTNCNPKDTCGPFVPHSNTLKRKTIQ 660
661 ILGQEPDAEMVEYLI 675
661 ILGQEPDAEMVEYLI 675
ESTL 2
8VHG2 PRELIMINARY; PRT; 891 AA.
C Q8VHG2;
C Q8VHG2;
T 01-MAR-2002 (T-EMBLrel. 20, Created)
T 01-MAR-2002 (T-EMBLrel. 20, Last sequence update)
T 01-OCT-2003 (T-EMBLrel. 25, Last annotation update)
E Angiomotin.
N AMOT.
S Mus musculus (Mouse).
C Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
C Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
X NCBI_TaxID=10090;
N [1]
P SEQUENCE FROM N.A.
C STRAIN=B6CBAF2; TISSUE=Placenta;
A Troyanovsky B., Bratt A., Holmgren L.;
L Submitted (DEC-2001) to the EMBL/GenBank/DBJ databases.
R EMBL; AF461135; AAL73436.1; -.
R MGD; MGI:108440; Amot.
Q SEQUENCE 891 AA; 95111 MW; C0544542A4CD6088 CRC64;
Query Match 80.2%; Score 2683; DB 11; Length 891;

```

```

Best Local Similarity 76.8%; Pred. No. 5.5e-109;
Matches 573; Conservative 29; Mismatches 62; Indels 82; Gaps 11;
QY 1 MPRAOPSSASQVPADPFAIVSRAQQWVEILSDENRLROELGCGYKVARLQKVTEI 60
DB 157 MPRAQ-QSASYQPMADPFAIVSRAQQWVEILSDENRLROELGCGYKVARLQKVTEI 215
QY 61 QVSVAYENLVKSSSKREALSKAMENKLEGEIRRMHDFNRDLRERLETANKQLAEKEYEG 120
DB 216 QVSVAYENLVKSSSKREALSKAMENKLEGEIRRMHDFNRDLRERLETANKQLAEKEYEG 275
QY 121 SDTTRKTTISQLPACKKESQREKELEALATARSNEDORRHIERDOALSNAQ 174
DB 276 SDTTRKTTISQLPACKKESQREKELEALATARSNEDORRHIERDOALSNAQ 329
QY 175 AKVVKLEELKKKKQVYVDKVKMOQALVOLQAACEKEQLEHRLRTRLELESRLRQ 234
DB 330 AK-WKLEELKKKKQVYVDKVKMOQALVOLQAACEKEQLEHRLRTRLELESRLRQ 388
QY 235 QNCQPTNVSEYNAALMELLREKEERILALEADMTKWEQKYLEENVMRHFALDAATVA 294
DB 389 QNCQPTNVSEYNAALMELLREKEERILALEADMTKWEQKYLEENVMRHFALDAATVA 448
QY 295 AORDTTVLSHSPNTSYDTALEARIQKEEBEILMANKRCLDMGRIKTLHAQIIEKDMIK 354
DB 449 AORDTTVLSHSPNTSYDTALEARIQKEEBEILMANKRCLDMGRIKTLHAQIIEKDMIK 508
QY 355 VIQORSRKE-RETEQLSSMRPAKSLMSISNAGSGLLSHSSTLTGSPIMEEKRDDKSKWGS 414
DB 509 VIQORSRKE-RETEQLSSMRPAKSLMSISNAGSGLLSHSSTLTGSPIMEEKRDDKSKWGS 567
QY 415 LGILGGDYRAEYVSTPSPVPPTPLLSAHSKTSGRDCSTQTERGTSNKTAAPVAPISV 474
DB 568 LGILGGDYRAEYVSTPSPVPPTPLLSAHSKTSGRDCSTQTERGTSNKTAAPVAPISV 627
QY 475 P--APVAAAATAAAI-----TATAATITTTWVAAPVAVAAAAAPAAAAA--PSPTAA 524
DB 628 PMAGPVAAAATAAAI-----TATAATITTTWVAAPVAVAAAAAPAAAAA--PSPTAA 687
QY 525 ATAAA-----VSPAAGQIIPAAASV----- 544
DB 688 ALAAAAAATPATSVSAAATSVSAANSISPAAPVAPVAPVAPVAPVAPVAPVAPVAPVAPV 747
QY 545 -----ASAAAVAPSAASAAAAVQVAPAPVAPVAPVAPVAPVAPVAPVAPVAPVAPV 589
DB 748 VSPATAATAAAVAAATTAATAATTAATAATTAATAATTAATAATTAATAATTAATAATTA 805
QY 590 AOTQAPTAPVAPVAPVAPVAPVAPVAPVAPVAPVAPVAPVAPVAPVAPVAPVAPVAPV 649
DB 806 TPTQASTPAPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPT 675
QY 650 SNTLKRKTIQILGQEPDAEMVEYLI 675
DB 866 SNTLKRKTIQILGQEPDAEMVEYLI 891
RESULT 3
Q8IY63 PRELIMINARY; PRT; 956 AA.
ID Q8IY63
AC Q8IY63;
DT 01-MAR-2003 (T-EMBLrel. 23, Created)
DT 01-MAR-2003 (T-EMBLrel. 23, Last sequence update)
DT 01-MAR-2003 (T-EMBLrel. 23, Last annotation update)
DE Similar to angiomotin like 1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Muscle;
RA Strausberg R.;
RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.

```

EMBL: BC037539; AAH37539.1; -- 6F47AE3A13816B95 CRC64;
 SEQUENCE 956 AA; 106574 MW; 647AE3A13816B95 CRC64;
 Query Match 43.3%; Score 1448; DB 4; Length 956;
 Best Local Similarity 62.0%; Pred. No. 2.7e-55;
 Matches 292; Conservative 83; Mismatches 82; Indels 14; Gaps 6;
 2 PRAOPSSAYQVPADPPAIVSRACQWVEILSDNRLRQELCYEKVAKVQVETEIQ 61
 418 PQQPPAASPQSLGPDADPAIVRAQWVEILTEENRVLHQLQYNDRLKHFKEIQ 477
 62 RVSEAYENLVKSSKREALEKAMNKLEGEIRRMHDFNRDLRERLETANKOLAEKEYEGS 121
 478 RISEAYESLVKSTTKGSLDKAMNKLEGEIRRLHDFNRDLRERLETANKOLSSREYEGH 537
 122 EDTRKTIQSLFAKNKESOREKEKLEALATARSNEDORRHIEIRDQALNAQAKVYLE 181
 538 ED-KAAEGHYASQNKFEKLEKLEMLAAVRTASEDHRRHIEILDQALNAQAKVYLE 596
 182 BELKKQYVDKVEKMQQALVQLAQACEKREQLBRLRLTRLELESRLIQRGNCQPT 241
 597 EELREKQAYVEKVEKQALQALQALQALQALQALQALQALQALQALQALQALQALQ 656
 242 NVSEYNAALMELLREKEERILALADMTKWEQKYLEENVRHPEALDAATVAQRDTTV 301
 657 NNPEYNAALMELLREKEERILALADMTKWEQKYLEENVRHPEALDAATVAQRDTTV 716
 302 ISHSNTSY-DTALREARIQKEEELMANKKCLDMEGRITKTHAQLIEKDMIKVLQORS 360
 717 INHSRNGSYGSSLEAHLWQEEVEVQANRCQDMVEYTKNHLAKIEKDMIKVLQORS 776
 361 RKEPKTSQSCMRPAKSLMSISNAGSLLSHTLGTGSPIMEKRDCKSWKSGILG 420
 777 RKDAGKTDSSSLRPARSVPSIA-AATGTHSRQTSLSSTQALAEKKEKTKWKSIGILG 834
 421 GDYRAEYVPSPTSPVPPTSP-----LSAHSKTSGRDCSTQTERGT 462
 835 KEHH-EHASAPLLPPPTPSALSSTASITTAASAAKTSKDSSTQTKSAE 884
 RESULT 4
 ID4H4 PRELIMINARY; PRT; 882 AA.
 Q9D4H4
 Q9D4H4
 01-JUN-2001 (TREMELrel. 17, Created)
 01-JUN-2001 (TREMELrel. 17, Last sequence update)
 01-OCT-2003 (TREMELrel. 25, Last annotation update)
 4932416D09Rik protein.
 AW07L1 OR 4932416D09Rik.
 Mus musculus (Mouse).
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 [1]
 SEQUENCE FROM N.A.
 STRAIN=C57BL/6J; TISSUE=Testis;
 MEDLINE=21085660; PubMed=11217851;
 Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Iehi Y.,
 Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
 Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaoka I.,
 Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
 Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
 Pletschmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
 Kuehl P., Lewis S., Matsuo Y., Nikaide I., Pesole G., Quackenbush J.,
 Schriml L.M., Staubli F., Suzuki R., Tonita M., Wagner L., Washio T.,
 Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
 Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
 Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
 Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
 Lyons P., Marchionni L., Mashima J., Mazzarelli J., Wombauts P.,
 Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
 Sasaki H., Sato K., Schoenbach C., Seta T., Shibata Y., Storch K.-F.,
 Suzuki H., Sutoy-oka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,

Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohsaki S.,
 RA Hayashizaki Y.;
 RT "Functional annotation of a full-length mouse cDNA collection.";
 RL Nature 409:685-690(2001).
 DR EMBL; AK016526; BAB30287.1; --
 DR MGD; MGI:1922973; Anot11.
 DR GO; GO:0005923; C:tight junction; IDA.
 SQ SEQUENCE 882 AA; 98423 MW; 88AAFE83FFA56046 CRC64;
 Query Match 41.8%; Score 1400; DB 11; Length 882;
 Best Local Similarity 58.2%; Pred. No. 3e-53;
 Matches 299; Conservative 80; Mismatches 103; Indels 32; Gaps 11;
 QY 1 MPRAOP-SSAYQVPADP-----FAIVSRACQWVEILSDNRLRQELCYEKVA 51
 DB 333 LPLPLPISLAASQPLPASPNOQLGPDADPAIVRAQWVEILTEENRVLHQLQYND 392
 QY 52 RLOKVETIQVSRAYENLVKSSKREALEKAMNKLEGEIRRMHDFNRDLRERLETANK 111
 DB 393 KLHKEKELQSISEAYESLVKSTTKGSLDKAMNKLEGEIRRLHDFNRDLRERLETAN 452
 QY 112 QLAEEKEYEGSEDTRKTIQSLFAKNKESOREKEKLEALATARSNEDORRHIEIRDQALS 171
 DB 453 QLSSEYDGHED-KAAESHVYSQNKFEKLEKLEMLAAVRTASEDHRRHIEILDQALS 511
 QY 172 NQAKVVKLEHLKKQYVDKVEKMQQALVQLAQACEKREQLBRLRLTRLELESRLI 231
 DB 512 NQAKVVKLEHLKKQYVDKVEKMQQALVQLAQACEKREQLBRLRLTRLELESRLI 571
 QY 232 QORQNCQPTVSEYNAALMELLREKEERILALADMTKWEQKYLEENVRHPEALDAAA 291
 DB 572 QKHGTGTPVPSIPECNAPALMELVREKEERILALADMTKWEQKYLEESTIRHFAMSA 631
 QY 292 TVAARDTIVTSHSNTSY-DTALREARIQKEEELMANKKCLDMEGRITKTHAQLIEK 350
 DB 632 AATAERDITINHSRNGSYGSSLEAHLWQEEVEVQANRCQDMVEYTKNHLAKIEK 691
 QY 351 AMIKVLQORSRKPSKTEQLSCMRPAKSLMSISNAGSLLSHTLGTGSPIMEKRDCK- 409
 DB 692 AMIKVLQORSRKDAGKTDSSS-LRPARSVPSIA-AATGTHSRQTSLSSTQALAEKKE 749
 QY 410 SWKSGILG-----GDYRAEYVPSPTSP---PVPSTPELLSA-HSKTGRDCSTQTERGT 461
 DB 750 TWKSGISGLGKHEHQOQASAPLLPTTPASALSLEPASITTSASSTHAKTSGKDSSTQTKST 809
 QY 462 E-----SNKTAAPVAPISVPAPVAAATAA 485
 DB 810 ELFWPSMASLPSRGRSLSTAPSNSPILKHPAAKGA 843
 RESULT 5
 Q96CM5 PRELIMINARY; PRT; 498 AA.
 ID Q96CM5
 AC Q96CM5
 DT 01-DEC-2001 (TREMELrel. 19, Created)
 DT 01-DEC-2001 (TREMELrel. 19, Last sequence update)
 DT 01-DEC-2001 (TREMELrel. 19, Last annotation update)
 DE Similar to RIKEN cDNA 4932416D09 gene (Fragment).
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Skin;
 RA Strausberg R.;
 RL Submitted (SEP-2001) to the EMBL/GenBank/DBSJ databases.
 DR EMBL; BC014126; AAH4126.1; --
 FT NON_TER 1
 SQ SEQUENCE 498 AA; 56333 MW; AD557A70AB2EDD44 CRC64;
 Query Match 40.0%; Score 1338; DB 4; Length 498;
 Best Local Similarity 62.7%; Pred. No. 7.9e-51;

Matches 269; Conservative 75; Mismatches 71; Indels 14; Gaps 6;

44 EGYCYKVARLOKVEIQRVSAYENLVKSSKREALEKAMNKLEGEIRWMDPFDLR 103
 2 EGYVNDKHLKFEKELORISAYESLVKSTTKESLDKAMNKLEGEIRLHDFNEDLR 61
 104 ELETANKOLAEKEVEGGEDTRKTIISQIFAKNKBSQREKELEAEATARSTNEDQRRHI 163
 62 DLELTANROLSPREVEGHED-KAASGHYASQNKFLKEKEKLEMLELAAVRTASEDHRRHI 120
 164 ETRDQALSNAQKVKLEBEELKKQVYVDKVKQQAALVOLQAACEKREBLEHRLTRLS 223
 121 EILDQALSNAQKVKLEBEELKKQVYVDKVKQQAALVOLQAACEKREBLEHRLTRLE 180
 224 ELESIRIOROGNOGNOPTNVSEYNAAALMELLREKBERILALADMTKWKQKYLEENVMR 283
 181 RELDALTUQKNGGONPAMPYFNAPALLELVREKEERILALEADMTKWKQKYLEESTIR 240
 284 HPALDAAATVAQORDTTVISHSPNTSY-DTALARIQKEEBEILMANKRCCLDMGRKTL 342
 241 HPANAAATAAABERDTTIIHNSRNGSYGESSLEAHWQEEVEVQANRCQDMETTKNL 300
 343 HAQILEKDAWIKVLOORSKPKESKTEQLSCMRPAKSLMSISNAGSGLLSHSTLTGSPIM 402
 301 HAKIIEKDAWIKVLOORSKPKESKTEQLSCMRPAKSLMSISNAGSGLLSHSTLTGSPIM 358
 403 BEKRDKSKWKSIGILLGGDYRAEYVPTSPVPPSTEL-----LSAHSKTSGRDC 453
 359 ESKKEETKWSIGILLGKHEH-EHASAPLLPPPTTSALSSIASTTAASSAHAKTGSNDS 417
 454 STQTERGTE 462
 418 STQTDKSAE 426

RESULT 6
 8K371 PRELIMINARY; PRT; 772 AA.

Q8K371
 01-OCT-2002 (T-EMBLrel. 22, Created)
 01-OCT-2002 (T-EMBLrel. 22, Last sequence update)
 01-OCT-2003 (T-EMBLrel. 25, Last annotation update)
 Similar to angiotensin like 2.
 ANOTL2.
 Mus musculus (Mouse).
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 NCBI_TaxID=10090;
 [1]
 SEQUENCE FROM N.A.
 Strausberg R.;
 Submitted (APR-2002) to the EMBL/GenBank/DBJ databases.
 EMBL; BC027824; AAH27824.1; -.
 MGD; MG1:1929286; Amotl2.
 SQ SEQUENCE 772 AA; 85278 MW; 997995F4D2A54989 CRC64;

Query Match 32.6%; Score 1090.5; DB 11; Length 772;
 Best Local Similarity 50.4%; Pred. No. 7.1e-40;
 Matches 244; Conservative 71; Mismatches 118; Indels 51; Gaps 9;

6 PSSASYQVPADPPAIV-----SRAQQWVEILSD-----ENRNLORELESCYKVARLOK 55
 284 PSSFGPVAVEGPPSAQATLGAHIAQVETLRNARLQDRNRELORELESTSEKAGRIEK 343
 56 VETRIQVSEAVENLVKSSKREALEKAMNKLEGEIRRMHDFNRDLRELETANKQLAE 115
 344 LENEIQRLSEAHESLMTSSKREALEKTMKNKMDGEMRLQDFNRDLRELESANRHLAS 403
 116 KEYEGSEDTKTIISQIFAKNKBSQREKELEAEATARSTNEDQRRHIETRDQALSNAQ 175
 404 KTQEAQAGSQDMWAKLLAQSYEQOQEQELEREMALLRGATIEDQRHAELEALQAGNQS 463
 176 KVKVLEBEELKKQVYVDKVKQQAALVOLQAACEKREBLEHRLTRLELESIRIQORQ 235

Db 464 RAARAEEELRKQAVYVVERLOQALQQAACEKREBLEHRLTRLEBLEKALRAQORQ 523
 QY 236 -----GNCQPTNVSEYNAAALMELLREKBERILALADMTKWKQKYLEENVMRHFALDA 289
 Db 524 TOTLAGGSGSGGSHLSALRLSQLEKEBEQILALEADMTKWKQKYLEENVMRHFALDA 583
 QY 290 AATVAQORDTTVISHSPNTSYDTALARIQKEEBEILMANKRCCLDMGRKTLHAQILEK 349
 Db 584 AATAAQAQRTTLIRHSPQSPSSSP-----NEGLIPGNHRHQEKESRLKVLHAQILEK 636
 QY 350 DAMIKVLOORSKPKESKTEQLSCMRPAKSLMSISNAGSGLLSHSTLTGSPIMEKRDCK 409
 Db 637 DAVIKVLOORSKPKESKTEQLSCMRPAKSLMSISNAGSGLLSHSTLTGSPIMEKRDCK 466
 QY 410 SWKSLGILLGGDYRAEYVPTSPVPPSTPLSAHSKTSGRDCSTQTERGTEBNT 466
 Db 677 GWQG-----LVSSERQTDARPDAGDRVPAEPEPATAPLPAHTKHGSRDGTQD-GPADNTS 731
 QY 467 AAVA 470
 Db 732 ACIA 735

RESULT 7
 Q8BS08 PRELIMINARY; PRT; 463 AA.

AC Q8BS08
 01-MAR-2003 (T-EMBLrel. 23, Created)
 01-MAR-2003 (T-EMBLrel. 23, Last sequence update)
 01-OCT-2003 (T-EMBLrel. 25, Last annotation update)
 Leman coiled-coil protein.
 ANOTL2.
 Mus musculus (Mouse).
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 NCBI_TaxID=10090;
 [1]
 SEQUENCE FROM N.A.
 STRAIN=C57BL/6J; TISSUE=Aorta, and Vein;
 MEDLINE=22354683; PubMed=12466851;
 RA The FANTOM Consortium.
 RA the RIKEN Genome Exploration Research Group Phase I & II Team;
 RT "Analysis of the mouse transcriptome based on functional annotation of
 60,770 full-length cDNAs.";
 RL Nature 420:563-573(2002).
 DR EMBL; AK040812; BAC30740.1; -.
 DR PIR; PT0534; PT0534.
 DR MGD; MG1:1929286; Amotl2.
 SQ SEQUENCE 463 AA; 51758 MW; 2F7BES2FBC8B1678 CRC64;

Query Match 32.5%; Score 1087.5; DB 11; Length 463;
 Best Local Similarity 52.7%; Pred. No. 5.6e-40;
 Matches 236; Conservative 68; Mismatches 103; Indels 41; Gaps 7;

QY 32 LSENENLORELESCYKVARLOKVEIQRVSAYENLVKSSKREALEKAMNKLEGE 91
 Db 11 LQDRNRELORELESTSEKAGRIEKLENEIQRLSEAHESLMTSSKREALEKTMKNKMDGE 70
 QY 92 IRRMHDNFRDLRELETANKQLAEKEYEGSEDTKTIISQIFAKNKBSQREKELEAEAT 151
 Db 71 MRQLQDFNRDLRELESANRHLASKTQEAQAGSQDMWAKLLAQSYEQOQEQEKLREEMAL 130
 QY 152 ARSTNEQRRHIEIRDOALSNAQKVKVLEBEELKKQVYVDKVKQQAALVOLQAACEKR 211
 Db 131 LRGAIEDQRHAELEALQAGNQSRAARAEEELRKQAVYVVERLOQALQQAACEKR 190
 QY 212 BOLHRLTRLELESIRIQORQ-----GNCQPTNVSEYNAAALMELLREKBERILAL 265
 Db 191 BOLHRLTRLELESIRIQORQ-----GNCQPTNVSEYNAAALMELLREKBERILAL 250
 QY 266 BADMTKWKQKYLEENVMRHFALDAAATVAQORDTTVISHSPNTSYDTALARIQKEEBE 325


```
32 LSDNRNLROBLECYEKVARLOKVEIQRVSEAVENLVKSSKEALEKAMRNKLEGE 91
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
11 LQDNRLQRELESSAKAGIKLESEIQLSEAHESLTRASKEALEKTMKNKNDSE 70
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
92 IRMVDNRDLRERLETANKOLAKVEYEGSEDTRKTIISQLFAKNKVSQBEKLEBLAT 151
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
71 MRLLQDNRLRERLESANRLASKTQEAQAGQDMVAKLLAQSYYQQQBEKLEBEMAL 130
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
152 ARSTNEDORRHIEIRDQALNSAOKAVVKLSEELKKQVYVDKVKMQQALVOLQAACEK 211
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
131 LRGAIEORRAELLEQALGNAQGRARAABEELKKQAYVEKVERIQQALQOLQAACEK 190
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
212 EQLERLRLERLESRLIQORQ-----GNCQPTNVSEYNAALMELLREKEBERILALE 266
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
191 EQLERLRLERLEQALKALRAQORQAGPGSGSGGSPELSRLSEQLREKEBEQILALE 250
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
267 ADMTKWQKYLEENVMRHPALDAATVAQORDTTVLISHSPNTSYDTALEARLQKBEETIL 326
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
251 ADMTKWQKYLEENVMRHPALDAATVAQORDTTVLISHSPNTSYDTALEARLQKBEETIL 303
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
327 MANRCLDMEGRIKTLHAQIIEKDAMIKVLQORSRKPSKTEQLSCMRPAKSLMSISNAG 386
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
304 TGHRHQEMESLKVLAHQILEKDAVILKVLQORSRDPGKAIQGS-LRPAKSPVSVFAAA 362
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
387 SGLSHSSTLTGSPIMEEKRDKSKWKSGLIGLGGDYRAEYVPS-----TPSPVPS 438
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
363 AA-----GTQWQOG-----LSSERQTDAPARLTDRAPTEEPVTA 400
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
439 TPLLSAHSKTSGRDCSTQTERGTESNKTAAVAPISVPAP-----VAAATAAATAAT 493
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
401 PP-AAHAKHSGRDSGTQTEGPPDSTSTCL-----PEPDSLLGCSSQRAASLDVATS 453
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 13
Y2J4
Q9V2J4 PRELIMINARY; PRT; 859 AA.
01-NOV-1999 (TRENBLrel. 12, Created)
01-NOV-1999 (TRENBLrel. 12, Last sequence update)
01-NOV-2003 (TRENBLrel. 23, Last annotation update)
Hypothetical protein KIAA0989 (Fragment).
KIAA0989.
Homo sapiens (Human).
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
NCBI_TaxID=9606;
[1]
SEQUENCE FROM N.A.
TISSUE=Brain;
MEDLINE=99246063; PubMed=10231032;
Nagase T., Ishikawa K., Suyama M., Kikuno R., Hirose M.,
Miyajima N., Tanaka A., Kohani H., Nomura N., Ohara O.,
"Prediction of the coding sequences of unidentified human genes. XIII.
The complete sequences of 100 new cDNA clones from brain which code
for large proteins in vitro.",
DNA Res. 6:63-70(1999).
EMBL; AB023206; BAA76833.1; -.
Genew; HGNC:17812; AMOTL2.
InterPro; IPR00408; Reg chr condens.
PROSITE; PS00626; RCCL_2; 1.
Hypothetical protein.
NON_TER 1
SEQUENCE 859 AA; 94325 MW; 5AF7E211E4CF73BD CRC64;

Query Match 31.1%; Score 1040.5; DB 4; Length 859;
Best Local Similarity 46.6%; Pred. No. 1:2e-37;
Matches 237; Conservative 80; Mismatches 127; Indels 65; Gaps 10;

13 PVPDPFAIVSRAQQWVEI-----LSDNRLNRLQELCGCYEKVARLOKVEIQRV 63
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
375 PVSAQASSATSGSAHLAQMEAVLRNARLQDNRLQRELESSAKAGRIKLESITORL 434
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 14
Q8TER8 PRELIMINARY; PRT; 402 AA.
AC Q8TER8;
DT 01-JUN-2002 (TRENBLrel. 21, Created)
DT 01-JUN-2002 (TRENBLrel. 21, Last sequence update)
DT 01-OCT-2002 (TRENBLrel. 22, Last annotation update)
DE Hypothetical protein (Fragment).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
NCBI_TaxID=9606;
[1]
SEQUENCE FROM N.A.
TISSUE=Pancreas;
RA Strausberg R.;
RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC025981; AAH25981.1; -.
KW Hypothetical protein.
FT NON_TER 1
SQ SEQUENCE 402 AA; 44234 MW; 5AD4ADF0E8A83676 CRC64;

Query Match 25.9%; Score 867.5; DB 4; Length 402;
Best Local Similarity 47.3%; Pred. No. 1:7e-30;
Matches 200; Conservative 62; Mismatches 106; Indels 55; Gaps 9;

89 EGEIRRMHDFNRDLRERLETANKOLAKVEYEGSEDTRKTIISQLFAKNKVSQBEKLEAE 148
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
4 DSEMRLODFNRDLRERLESANRLASKTQEAQAGSQDMVAKLLAQSYYQQQBEKLE 63
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
149 LATASTVEDORRHIEIRDQALNSAOKAVVKLSEELKKQVYVDKVKMQQALVOLQAAC 208
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
64 MALLRGAEIQDORRAELLEQALGNAQGRARAABEELKKQAYVEKVERIQQALQOLQAAC 123
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
209 EKEQLEHLRLERLESRLIQORQ-----GNCQPTNVSEYNAALMELLREKEBERIL 263
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
124 EKEQLEHLRLERLEQALKALRAQORQAGPGSGSGGSPELSRLSEQLREKEBEQIL 183
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
264 ALEADMTKQKYLEENVMRHPALDAATVAQORDTTVLISHSPNTSYDTALEARLQKBE 323
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
```

earch completed: June 16, 2004, 19:14:52
op time : 65.7139 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

1 protein - protein search, using sw model

on: June 16, 2004, 19:06:57 ; Search time 17.8313 Seconds

(without alignments)
2365.920 Million cell updates/sec

itle: US-09-332-063-4

irect score: 663

quence: 1 BSNKTAAPVAPVAPVAAA.....QASAPATQAPTAPVAPT 143

oring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

arched: 1586107 seqs, 282547505 residues

otal number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

ost-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

atabase : A_Geneseq_29Jan04:*
1: Genesep1980s:*
2: Genesep1990s:*
3: Genesep2000s:*
4: Genesep2001s:*
5: Genesep2002s:*
6: Genesep2003as:*
7: Genesep2003bs:*
8: Genesep2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

result No.	Score	Query Match	Length	DB ID	Description
1	663	100.0	143	3	AAY54054 Angiostat
2	663	100.0	479	5	ABP43965 Unidentif
3	663	100.0	675	3	AAY54053 A variant
4	663	100.0	675	3	AAY54052 An angiot
5	647	97.6	993	7	AAB37918 Human CGD
6	226	34.1	354	6	ABU21693 Protein e
7	218.5	33.0	309	7	ABO23523 Pseudomon
8	218.5	32.9	103	5	ABP78540 Ala-Pro-A
9	215.5	32.5	2701	6	ABR92087 Human cer
10	215.5	32.5	2819	4	AAB35408 Human Q7C
11	209	31.5	1259	6	ABR57561 Human MC2
12	205.5	31.0	842	4	ABR66831 Drosophil
13	205.5	31.0	864	4	ABR71319 Drosophil
14	205.5	31.0	1743	6	ABU8255 Novel hum
15	205.5	31.0	1743	6	ABU90134 Novel hum
16	205.5	31.0	1743	6	ABU96436 Novel hum
17	205.5	31.0	1743	6	ABU99045 Novel hum
18	205.5	31.0	1743	6	ABU98260 Novel hum
19	205.5	31.0	1743	6	ABU91966 Novel hum
20	205.5	31.0	1743	6	ABU85270 Novel hum
21	205.5	31.0	1743	6	ABO00409 Novel hum
22	205.5	31.0	1743	6	ABU88960 Novel hum
23	205.5	31.0	1743	6	ABO06456 Novel hum
24	205.5	31.0	1743	6	ABU95516 Novel hum
25	205.5	31.0	1743	6	ABU95206 Novel hum

ALIGNMENTS

RESULT 1

AAY54054
ID AAY54054 standard; protein; 143 AA.

XX
AC AAY54054;
XX
XX
DT 27-MAR-2000 (first entry)
XX
DS Angiostatin-binding domain of ABP-1, designated Big-3.
XX
KW Human; angiogenesis-associated protein; plasminogen; ABP-1;
KW kringle domain; angiotatin; plasminogen receptor;
KW angiogenesis-related disease; tumor; diabetes; rheumatoid arthritis;
KW inflammatory disease; psoriasis; chronic inflammation; intestine; asthma;
KW obesity; gene therapy; angiotatin-binding domain; Big-3.
XX
OS Homo sapiens.
XX
PN MO9966038-A1.
XX
PD 23-DEC-1999.
XX
PF 11-JUN-1999; 99WO-EP004109.
XX
PR 15-JUN-1998; 98SE-00002130.
PR 15-JUN-1998; 98US-0089266P.
PR 17-DEC-1998; 98SE-00004372.
PR 23-DEC-1998; 98US-0114386P.
XX
PA (PHAA) PHARMACIA & UPJOHN AB.
XX
PI Holmgren L, Troyanovsky B;
XX
DR WPI; 2000-106099/09.
PT Novel human protein useful for treating angiogenesis associated diseases or disorders.
XX
XX Claim 6; Page 49; 58pp; English.
XX
CC The present sequence represents the angiotatin-binding domain, designated Big-3, of a human angiogenesis-associated protein. The protein binds an N-terminal fragment of plasminogen. The ABP-1 protein is designated ABP-1, and binds the first 4 kringle domains (K1-K4) and/or kringle 5 (K5) of plasminogen. These four kringle domains comprise angiotatin. The ABP-1 protein acts as a receptor for plasminogen. A polymorphic variant of ABP-1 is also described, in AAY54053. ABP-1 can be used to manufacture medicaments for treating angiogenesis-related

ABU90754 Novel hum
ABU93916 Novel hum
ABU86190 Novel hum
ABU82045 Novel hum
ABU07906 Novel hum
ABU94226 Novel hum
ABO00099 Novel hum
ABU87110 Novel hum
ABU91351 Novel hum
ABU90444 Novel hum
ABU97035 Novel hum
ABO05231 Novel hum
ABU3775 Protein e
ABU71039 Drosophil
ABG04954 Novel hum
AAW26540 Trypanoso
AAY23307 Trypanoso
ABG20119 Novel hum
ABU59024 Drosophil
AAW06913 T. cruzi

26 205.5 31.0 1743 6 ABU90754
27 205.5 31.0 1743 6 ABU93916
28 205.5 31.0 1743 6 ABU86190
29 205.5 31.0 1743 6 ABU82045
30 205.5 31.0 1743 6 ABU07906
31 205.5 31.0 1743 6 ABU94226
32 205.5 31.0 1743 6 ABO00099
33 205.5 31.0 1743 6 ABU87110
34 205.5 31.0 1743 6 ABU91351
35 205.5 31.0 1743 6 ABU90444
36 205.5 31.0 1743 6 ABU97035
37 205.5 31.0 1743 6 ABO05231
38 205 30.9 329 6 ABU3775
39 204.5 30.8 1013 4 ABU71039
40 201.5 30.4 806 4 ABG04954
41 200.5 30.2 442 2 AAW26540
42 200.5 30.2 442 2 AAY23307
43 200.5 30.2 2724 4 ABG20119
44 200 30.2 477 4 ABU59024
45 197 29.7 262 2 AAW06913

The present sequence represents a polymorphic variant of a human angiogenesis-associated protein which is able to bind an N-terminal fragment of plasminogen. The native protein, described in AAY54052, is designated ABP-1, and binds the first 4 kringle domains (K1-K4) and/or kringle 5 (K5) of plasminogen. These four kringle domains comprise angiotensin. The ABP-1 protein acts as a receptor for plasminogen. The angiotensin-binding domain of the ABP-1 protein is described in AAY54054. ABP-1 can be used to manufacture medicaments for treating angiogenesis-related diseases or disorders, such as tumor conditions, diabetes, rheumatoid arthritis, and even some inflammatory diseases such as psoriasis, chronic inflammation of the intestine, asthma, etc. The ABP-1 protein may also be able to treat and cure, or prevent, obesity. The ABP-1 DNA can be used in gene therapy techniques

Sequence 675 AA;

Query Match 100.0%; Score 663; DB 3; Length 675;
Best Local Similarity 100.0%; Pred. No. 2.2e-38;
Matches 143; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
1 ESNKTAAPVAPISVPAPVAAAATAAATATATATTTTVAAPVAVAAAAAPAAAAAPSPA 60
462 ESNKTAAPVAPISVPAPVAAAATAAATATATTTTVAAPVAVAAAAAPAAAAAPSPA 521
61 TAAATAAASVPAAGQIPAAASVSAASAAVAPSAASAAAVQVAPAPVAPALVPVPAP 120
522 TAAATAAASVPAAGQIPAAASVSAASAAVAPSAASAAAVQVAPAPVAPALVPVPAP 581
121 AAAQASAPAQTOAPTSAPAVAPT 143
582 AAAQASAPAQTOAPTSAPAVAPT 604

RESULT 4
AAY54052
AAY54052 standard; protein; 675 AA.

AAY54052;

27-MAR-2000 (first entry)

An angiogenesis-associated protein which binds plasminogen.

Human; angiogenesis-associated protein; plasminogen; ABP-1;
kringle domain; angiotensin; plasminogen receptor;
angiogenesis-related disease; tumor; diabetes; rheumatoid arthritis;
inflammatory disease; psoriasis; chronic inflammation; intestine; asthma;
obesity; gene therapy.

Homo sapiens.

WO9966038-A1.

23-DEC-1999.

11-JUN-1999; 99WO-EP004109.

15-JUN-1998; 98SE-00002130.

15-JUN-1998; 98US-0089266P.

17-DEC-1998; 98SE-00004372.

29-DEC-1998; 98US-0114386P.

(PHAA) PHARMACIA & UPJOHN AB.

Holmgren L, Troyanovsky B;

WPI; 2000-106099/09.

N-PSDB; AA245329.

Novel human protein useful for treating angiogenesis associated diseases

or disorders.

Claim 4; Page 43-46; 58pp; English.

The present sequence represents a human angiogenesis-associated protein which is able to bind an N-terminal fragment of plasminogen. The protein is designated ABP-1, and binds the first 4 kringle domains (K1-K4) and/or kringle 5 (K5) of plasminogen. These four kringle domains comprise angiotensin. The protein acts as a receptor for plasminogen. The angiotensin-binding domain of the ABP-1 protein is described in AAY54054. A polymorphic variant of ABP-1 is also described, in AAY54053. ABP-1 can be used to manufacture medicaments for treating angiogenesis-related diseases or disorders, such as tumor conditions, diabetes, rheumatoid arthritis, and even some inflammatory diseases such as psoriasis, chronic inflammation of the intestine, asthma, etc. The protein may also be able to treat and cure, or prevent, obesity. The ABP-1 DNA can be used in gene therapy techniques

Sequence 675 AA;

Query Match 100.0%; Score 663; DB 3; Length 675;
Best Local Similarity 100.0%; Pred. No. 2.2e-38;
Matches 143; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
1 ESNKTAAPVAPISVPAPVAAAATAAATATATTTTVAAPVAVAAAAAPAAAAAPSPA 60
462 ESNKTAAPVAPISVPAPVAAAATAAATATATTTTVAAPVAVAAAAAPAAAAAPSPA 521
61 TAAATAAASVPAAGQIPAAASVSAASAAVAPSAASAAAVQVAPAPVAPALVPVPAP 120
522 TAAATAAASVPAAGQIPAAASVSAASAAVAPSAASAAAVQVAPAPVAPALVPVPAP 581
121 AAAQASAPAQTOAPTSAPAVAPT 143
582 AAAQASAPAQTOAPTSAPAVAPT 604

RESULT 5
AAE37918
ID AAE37918 standard; protein; 993 AA.

AAE37918;

06-NOV-2003 (first entry)

Human CGDD-7 protein.

Human; cell growth, differentiation and death protein; CGDD; leukaemia;
neurodegenerative disorder; Parkinson's disease; Alzheimer's disease;
muscular disorder; myotonic dystrophy; catatonia; endocrine disorder;
diabetes; Grave's disease; cancer; immunological disorder; scleroderma;
systemic lupus erythematosus; allergy; Crohn's disease; renal disorder;
gastrointestinal disorder; Goodpasture's syndrome; infection; cirrhosis;
cardiovascular disorder; atherosclerosis; hepatic disease; transgenic;
transgenic animal; gene therapy; neuroprotective; relaxant; cytostatic;
dermatological; immunosuppressive; cerebroprotective; anticonvulsant;
antibacterial; antiparasitic; fungicide; virucide; uropathic; cardiant;
protozoacide; nootropic.

Homo sapiens.

WO2003050253-A2.

19-JUN-2003.

04-DEC-2002; 2002WO-US039133.

07-DEC-2001; 2001US-0340747P.

20-DEC-2001; 2001US-034761P.

15-JAN-2002; 2002US-0349705P.

06-FEB-2002; 2002US-035476P.

12-FEB-2002; 2002US-0356216P.

(INCY-) INCYTE GENOMICS INC.

X ABO23523;
X 04-SSP-2003 (first entry)
X Pseudomonas aeruginosa outlier protein #3.
X Candidate protein identification; pathogen; anti-infective;
W outlier protein; virulence protein; antigen; drug target protein;
W pathogenic organism; antimicrobial.
X Pseudomonas aeruginosa.
S US2003039963-A1.
X 27-FEB-2003.
X 30-MAR-2001; 2001US-00820843.
X 30-MAR-2001; 2001US-00820843.
X (BRAH/) BRAHMACHARI S K.
X (RAMA/) RAMACHANDRAN S.
X (NAND/) NANDI T.
X (BHIM/) BHIMARAO C.
X Brahmachari SK, Ramachandran S, Nandi T, Bhimarao C;
X WPI; 2003-492159/46.
X Identifying candidate proteins useful as anti-infectives involves
X matching outlier protein sequences with protein sequences in databases.
X Example 7; Page 37-38; 117pp; English.
X The present invention relates to a method for identifying candidate
X proteins in pathogens useful as anti-infectives. The invention discloses
X a computational method which involves the calculation of several sequence
X attributes and their subsequent analysis results in the identification
X of outlier proteins in different pathogens. The method is useful for the
X identification of outlier proteins (e.g. virulence proteins, antigens or
X proteins used as drug targets) in pathogenic organisms. The method of the
X invention provides reproducible results as it does not depend on the
X variable biochemical characterisation of proteins. ABO23500-ABO23617
X represent outlier proteins identified from different pathogenic organisms
X
X Sequence 309 AA;
Query Match 33.0%; Score 218.5; DB 7; Length 309;
Best Local Similarity 45.0%; Pred. No. 7.2e-08;
Matches 63; Conservative 8; Mismatches 62; Indels 7; Gaps 3;
Y 4 KTAAPVAPISVPAPVAAATATAATATTTTVAAPVAVAAAAAPAAAAAPSPATAA 63
Y 159 KTAAPVAPVAPVAAATATAATATTTTVAAPVAVAAAAAPAAAAAPSPATAA 215
Y 64 ATAAVSPAAAGQIPAAASVAGAAVAPSAASAAVQVAPAA-PAPVPAPALVVPVAPAA 122
Y 216 AKAAPVAPVAPVAAATATAATATTTTVAAPVAVAAAAAPAAAAAPSPATAA 272
Y 123 AQASAPVAPVAPVAAATATAATATTTTVAAPVAVAAAAAPAAAAAPSPATAA 142
Y 273 AKPAPVAPVAPVAAATATAATATTTTVAAPVAVAAAAAPAAAAAPSPATAA 292
RESULT 8
BB78540
D ABB78540 standard; protein; 103 AA.
X ABB78540;
X 09-JUL-2002 (first entry)
X

DE Ala-Pro-Ala-Pro construct related protein sequence SEQ ID NO:240.
XX Plant; Gum arabic glycoprotein; GAGP; hydroxyproline-rich glycoprotein;
KW HRGP; repetitive proline-rich protein; RRRP; arabinogalactan protein;
KW AGP; plant gum.
XX Acacia senegal.
OS Synthetic.
XX WO200178503-A2.
XX 25-OCT-2001.
XX 12-APR-2001; 2001WO-US012336.
XX 12-APR-2000; 2000US-00547693.
XX (UYOH-) UNIV OHIO.
XX Kieliszewski MJ;
XX WPI; 2002-041307/05.
DR N-PSDB; ABL51829.
XX Nucleic acids and proteins useful for producing hydroxy-proline rich
PT glycoproteins in plants.
XX Example 25; Fig 18; 326pp; English.
XX The present invention describes synthetic genes encoding plant gums and
CC other hydroxyproline (Hyp)-rich glycoproteins (HRGPs) and the nucleic
CC acids that encode them. The nucleic acids, proteins and methods from the
CC present invention may be used to produce HRGPs, repetitive proline-rich
CC proteins (RRPs) and arabinogalactan-proteins (AGPs) in plants via
CC recombinant methodologies. Also described is the expression of synthetic
CC genes designed from repetitive peptide sequences, such as glycoproteins
CC (including the peptide sequences of gum arabic glycoprotein (GAGP)).
CC ABL51730 to ABL51849 and ABB78401 to ABB78544 represent sequences used in
CC the exemplification of the present invention
XX Sequence 103 AA;
Query Match 32.9%; Score 218; DB 5; Length 103;
Best Local Similarity 57.1%; Pred. No. 2.8e-08;
Matches 60; Conservative 3; Mismatches 36; Indels 6; Gaps 4;
OY 39 AAAPVAVAAAPVAPVAAATATAATATTTTVAAPVAVAAAAAPAAAAAPSPATAA 98
DB 1 APAP-AP 55
OY 99 VQVAPAP-APAPVAPALVVPVAPVAPVAPVAPVAPVAPVAPVAPVAPVAPVAP 142
DB 56 PAP 100
RESULT 9
ABR92087
ID ABR92087 standard; protein; 2701 AA.
XX ABR92087;
AC ABR92087;
XX 10-SEP-2003 (first entry)
DT Human cervical cancer cell marker encoding cDNA SEQ ID NO:83.
XX Human cervical cancer; cervical cancer marker; cancer therapy;
XX Human; cervical cancer; gene therapy; vaccine.
XX Homo sapiens.
XX WO2002101075-A2.
XX 19-DEC-2002.

X 12-JUN-2002; 2002WO-US018638.
 P
 R 13-JUN-2001; 2001US-029815SP.
 R 13-JUN-2001; 2001US-029815SP.
 R 14-NOV-2001; 2001US-0335936P.
 X (MILL-) MILLENNIUM PHARM INC.
 X Schlegel R, Chen Y, Zhao X, Monahan JE, Kamatkar S;
 I Gannavarapu M, Glatt K, Hoersch S;
 I WPI; 2003-156967/15.
 R N-PSDB; ACF12869.
 X
 X New isolated nucleic acid molecule useful for detecting, characterizing,
 T preventing and treating human cervical cancers, in various prognostic and
 T diagnostic assays, in pharmacogenomics and in monitoring clinical trials.
 X
 X Claim 4; Page 233-239; 386pp; English.
 X
 X ACF12828 to ACF12947 encode the human cervical cancer marker proteins (I)
 C given in ABR92047 to ABR92164. A higher level of expression of (I) than
 C normal indicates the presence of cervical cancer. Also described: (1) a
 C vector (II) containing (I); (2) a host cell (III) containing (I); and (3)
 C assessing (M1) whether a patient is afflicted with cervical cancer,
 C comprising comparing the level of expression of a marker in a patient's
 C sample, and the normal level of expression of the marker in the level of
 C -cervical cancer sample, where a significant increase in the level of
 C expression of the marker in the patient's sample relative to that in the
 C control sample is an indication that the patient is afflicted with
 C cervical cancer. (I) has cytostatic activity, and can be used in gene
 C therapy and in vaccines. (I) is useful in detecting, characterizing,
 C preventing and treating human cervical cancers. (II) may also be used in
 C various prognostic and diagnostic assays, pharmacogenomics and in
 C monitoring clinical trials
 X
 X Sequence 2701 AA;
 Q
 Query Match 32.5%; Score 215.5; DB 6; Length 2701;
 Best Local Similarity 34.1%; Pred. No. 8.8e-07;
 Matches 58; Conservative 25; Mismatches 48; Indels 39; Gaps 5;
 Y 5 TAAVAPISVPAPVAAAATA-----AAITATAATTTTVAAPVAAVAAAAPAAAAAP 57
 D 1781 TSAPVPASTLAPVLASTAPVPASPAPVSPASVSASVSPASTSAAITSSAPASAPAP 1840
 Y 58 SP-----ATAAATAAASVSPAAGQIPAAASVASAAVAPSAAA 95
 D 1841 TPILASVSTPASVTILASASIPILASALASTSAPTAPAASSPAPVPI--TAPTIPASAP 1898
 Y 96 AAAVQVAPA---APAPVPAPALVPVAPAAQAASAPAOQTAPTSAPAVAP 142
 D 1899 TASVPLAPASAPAPAPTVPSPAPNPAP-----PAPAQTOAQTHKPVQNP 1943
 RESULT 10
 IAB35408
 ID AAB35408 standard; protein; 2819 AA.
 C
 C AAB35408;
 C
 C 23-MAY-2001 (first entry)
 C
 C Human 07CG27 gene protein.
 C
 C Human; 07CG27 gene; chromosome 1; HPC1 region; prostate cancer; oncogene.
 C
 C Homo sapiens.
 C
 C WO200116291-A2.
 C
 C 08-MAR-2001.
 D

XX 25-AUG-2000; 2000WO-US023291.
 XX
 XX 27-AUG-1999; 99US-0151049P.
 XX
 XX (MYRI-) MYRIAD GENETICS INC.
 XX (HOSP-) HOSPITAL FOR SICK CHILDREN.
 PA
 XX Tavtigian SV, Swedlund B, Simard J, Rommens JM;
 PI
 XX WPI; 2001-226682/23.
 DR N-PSDB; AAF28060.
 DR
 XX Novel human prostate cancer marker gene termed as 07CG27 gene, useful for
 PT screening mutations in the gene in diagnosis of a predisposition to
 PT cancer.
 XX
 XX Claim 1; Page 91-99; 99pp; English.
 PS
 XX The present invention provides the protein and coding sequences of the
 CC human 07CG27 oncogene. This gene is found at the HPC1 region of
 CC chromosome 1. The sequences can be used in the diagnosis and
 CC identification of treatments for prostate cancer. The present sequence is
 CC the 07CG27 protein
 XX
 XX Sequence 2819 AA;
 SQ
 Query Match 32.5%; Score 215.5; DB 4; Length 2819;
 Best Local Similarity 34.1%; Pred. No. 9.2e-07;
 Matches 58; Conservative 25; Mismatches 48; Indels 39; Gaps 5;
 QY 5 TAAVAPISVPAPVAAAATA-----AAITATAATTTTVAAPVAAVAAAAPAAAAAP 57
 Db 1783 TSAPVPASTLAPVLASTAPVPASPAPVSPASVSASVSPASTSAAITSSAPASAPAP 1842
 QY 58 SP-----ATAAATAAASVSPAAGQIPAAASVASAAVAPSAAA 95
 Db 1843 TPILASVSTPASVTILASASIPILASALASTSAPTAPAASSPAPVPI--TAPTIPASAP 1900
 QY 96 AAAVQVAPA---APAPVPAPALVPVAPAAQAASAPAOQTAPTSAPAVAP 142
 Db 1901 TASVPLAPASAPAPAPTVPSPAPNPAP-----PAPAQTOAQTHKPVQNP 1945
 RESULT 11
 IABR57561
 ID ABR57561 standard; protein; 1259 AA.
 C
 C ABR57561;
 C
 C 04-AUG-2003 (first entry)
 C
 C Human MC25.
 C
 C Human; Antiallergic; Dermatological; Antiinflammatory; Gene therapy;
 C mast cell activation; allergic hypersensitivity; asthma; mastocytosis;
 C seasonal rhinitis; urticaria; atopic dermatitis.
 C
 C Homo sapiens.
 C
 C WO2003029464-A2.
 C
 C 10-APR-2003.
 C
 C 27-SEP-2002; 2002WO-EP010897.
 C
 C 01-OCT-2001; 2001US-0325536P.
 C
 C (UNIO) UCB SA.
 C
 C Nocka K, Medley Q, Thomas D, Gu J, Lu S;
 C
 C WPI; 2003-381640/36.
 D

PR	18-DEC-1997;	97US-0068017P;
PR	10-MAR-1998;	98US-0077450P;
PR	11-MAR-1998;	98US-0077632P;
PR	11-MAR-1998;	98US-0077649P;
PR	20-MAR-1998;	98US-0078886P;
PR	22-MAR-1998;	98US-0079539P;
PR	22-MAR-1998;	98US-0079549P;
PR	27-MAR-1998;	98US-0080178P;
PR	27-MAR-1998;	98US-0080110P;
PR	31-MAR-1998;	98US-0080194P;
PR	01-APR-1998;	98US-0080327P;
PR	01-APR-1998;	98US-0080333P;
PR	08-APR-1998;	98US-0081049P;
PR	08-APR-1998;	98US-0081070P;
PR	09-APR-1998;	98US-0081195P;
PR	09-APR-1998;	98US-0081938P;
PR	15-APR-1998;	98US-0082568P;
PR	21-APR-1998;	98US-0082569P;
PR	21-APR-1998;	98US-0082704P;
PR	22-APR-1998;	98US-0082732P;
PR	28-APR-1998;	98US-0083322P;
PR	29-APR-1998;	98US-0083495P;
PR	29-APR-1998;	98US-0083496P;
PR	29-APR-1998;	98US-0083499P;
PR	29-APR-1998;	98US-0083559P;
PR	05-MAY-1998;	98US-0084366P;
PR	06-MAY-1998;	98US-0084414P;
PR	07-MAY-1998;	98US-0084639P;
PR	07-MAY-1998;	98US-0084640P;
PR	07-MAY-1998;	98US-0085579P;
PR	15-MAY-1998;	98US-0085580P;
PR	15-MAY-1998;	98US-0085582P;
PR	15-MAY-1998;	98US-0085700P;
PR	18-MAY-1998;	98US-0086023P;
PR	22-MAY-1998;	98US-0086392P;
PR	22-MAY-1998;	98US-0086486P;
PR	28-MAY-1998;	98US-0087098P;
PR	28-MAY-1998;	98US-0087208P;
PR	02-JUN-1998;	98US-0087609P;
PR	02-JUN-1998;	98US-0087759P;
PR	03-JUN-1998;	98US-0087827P;
PR	04-JUN-1998;	98US-0088025P;
PR	04-JUN-1998;	98US-0088028P;
PR	04-JUN-1998;	98US-0088029P;
PR	04-JUN-1998;	98US-0088033P;
PR	04-JUN-1998;	98US-0088326P;
PR	05-JUN-1998;	98US-0088167P;
PR	05-JUN-1998;	98US-0088202P;
PR	05-JUN-1998;	98US-0088212P;
PR	05-JUN-1998;	98US-0088262P;
PR	09-JUN-1998;	98US-0088177P;
PR	09-JUN-1998;	98US-0088655P;
PR	10-JUN-1998;	98US-0088722P;
PR	10-JUN-1998;	98US-0088738P;
PR	10-JUN-1998;	98US-0088811P;
PR	10-JUN-1998;	98US-0088824P;
PR	10-JUN-1998;	98US-0088825P;
PR	10-JUN-1998;	98US-0088826P;
PR	11-JUN-1998;	98US-0088961P;
PR	11-JUN-1998;	98US-0088963P;
PR	11-JUN-1998;	98US-0088876P;
PR	12-JUN-1998;	98US-0089090P;
PR	12-JUN-1998;	98US-0089105P;
PR	16-JUN-1998;	98US-0089512P;
PR	16-JUN-1998;	98US-0089538P;
PR	17-JUN-1998;	98US-0089598P;
PR	17-JUN-1998;	98US-0089653P;
PR	18-JUN-1998;	98US-0089808P;
PR	19-JUN-1998;	98US-0089952P;
PR	22-JUN-1998;	98US-0090246P;
PR	22-JUN-1998;	98US-0090252P;

```

PR 22-JUN-1998; 98US-0090254P.
PR 24-JUN-1998; 98US-0090439P.
PR 24-JUN-1998; 98US-0090435P.
PR 24-JUN-1998; 98US-0090444P.
PR 24-JUN-1998; 98US-0090441P.
PR 24-JUN-1998; 98US-0090535P.
PR 24-JUN-1998; 98US-0090540P.
PR 25-JUN-1998; 98US-0090540P.
PR 25-JUN-1998; 98US-0090676P.
PR 25-JUN-1998; 98US-0090678P.
PR 25-JUN-1998; 98US-0090690P.
PR 25-JUN-1998; 98US-0090694P.
PR 25-JUN-1998; 98US-0090695P.
PR 25-JUN-1998; 98US-0090696P.
PR 26-JUN-1998; 98US-00105413.
PR 26-JUN-1998; 98US-0090862P.
PR 26-JUN-1998; 98US-0090863P.
PR 26-JUN-1998; 98US-0091010P.
PR 01-JUL-1998; 98US-0091359P.
PR 01-JUL-1998; 98US-0091544P.
PR 02-JUL-1998; 98US-0091478P.
PR 02-JUL-1998; 98US-0091486P.
PR 02-JUL-1998; 98US-0091626P.
PR 02-JUL-1998; 98US-0091628P.
PR 02-JUL-1998; 98US-0091632P.
PR 24-JUL-1998; 98US-0094006P.
PR 04-AUG-1998; 98US-0095282P.
PR 10-AUG-1998; 98US-0095998P.
PR 10-AUG-1998; 98US-0096012P.
PR 17-AUG-1998; 98US-0096757P.
PR 17-AUG-1998; 98US-0096766P.
PR 17-AUG-1998; 98US-0096867P.
PR 17-AUG-1998; 98US-0096891P.
PR 17-AUG-1998; 98US-0096897P.
PR 18-AUG-1998; 98US-0096949P.
PR 18-AUG-1998; 98US-0096959P.
PR 18-AUG-1998; 98US-0097022P.
PR 26-AUG-1998; 98US-0097952P.
PR 26-AUG-1998; 98US-0097952P.
PR 26-AUG-1998; 98US-0097955P.
PR 26-AUG-1998; 98US-0097955P.
PR 26-AUG-1998; 98US-0097971P.
PR 26-AUG-1998; 98US-0097974P.
PR 01-SEP-1998; 98US-0098014P.
PR 01-SEP-1998; 98US-0098716P.
PR 02-SEP-1998; 98US-0098723P.
PR 02-SEP-1998; 98US-0098803P.
PR 02-SEP-1998; 98US-0098821P.
PR 09-SEP-1998; 98US-0098843P.
PR 10-SEP-1998; 98US-0099602P.
PR 10-SEP-1998; 98US-0099741P.
PR 10-SEP-1998; 98US-0099754P.
PR 10-SEP-1998; 98US-0099763P.
PR 15-SEP-1998; 98US-0099812P.
PR 15-SEP-1998; 98US-0100388P.
PR 16-SEP-1998; 98US-0100662P.
PR 16-SEP-1998; 98US-0100664P.
PR 16-SEP-1998; 98US-0101751P.
PR 16-SEP-1998; 98MO-US019330.
PR 17-SEP-1998; 98US-0100683P.
PR 17-SEP-1998; 98US-0100684P.
PR 17-SEP-1998; 98US-0100915P.
PR 17-SEP-1998; 98US-0100930P.
PR 18-SEP-1998; 98US-0100949P.
PR 18-SEP-1998; 98US-0101014P.
PR 18-SEP-1998; 98US-0101068P.
PR 23-SEP-1998; 98US-0101471P.
PR 23-SEP-1998; 98US-0101472P.
PR 23-SEP-1998; 98US-0101475P.
PR 23-SEP-1998; 98US-0101477P.
PR 24-SEP-1998; 98US-0101738P.
PR 24-SEP-1998; 98US-0101739P.
PR 24-SEP-1998; 98US-0101743P.
PR 24-SEP-1998; 98US-0101922P.

PR 25-SEP-1998; 98US-0101786P.
PR 29-SEP-1998; 98US-0102207P.
PR 29-SEP-1998; 98US-0102240P.
PR 29-SEP-1998; 98US-0102330P.
PR 29-SEP-1998; 98US-0102331P.
PR 30-SEP-1998; 98US-0102487P.
PR 30-SEP-1998; 98US-0102570P.
PR 30-SEP-1998; 98US-0102571P.
PR 01-OCT-1998; 98US-0102684P.
PR 01-OCT-1998; 98US-0102687P.
PR 02-OCT-1998; 98US-0102965P.
PR 06-OCT-1998; 98US-0103258P.

Query Match 31.0%; Score 205.5; DB 6; Length 1743;
Best Local Similarity 57.6%; Pred. No. 2.9e-06;
Matches 57; Conservative 6; Mismatches 35; Indels 1; Gaps 1;

QY 5 TAAVAPISVPAPVAAAA TAAATATATATTTTAAAPVAVAAAAAPAAAAAPSPATAA 63
DB 1645 TAAACATTCATATAAACTAAATATATGATTCAAAAAATAAAAAAATAAAAAA 1704
QY 64 ATAAVSPAAAGQIPAAASVASAAVAPSAASAAVQVA 102
DB 1705 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1743

RESULT 15
ABU90134
ID ABU90134 standard; protein; 1743 AA.
AC ABU90134;
XX
DT 11-AUG-2003 (first entry)
DE
DE Novel human secreted and transmembrane PRO protein #3.
KW Human; gene therapy; tissue typing; tumour; chondrocyte proliferation;
KW chondrocyte differentiation; tumour necrosis factor-alpha release;
KW affinity purification.
XX
XX Homo sapiens.
XX
XX US2003036147-A1.
XX
XX 20-FEB-2003.
XX
XX 02-JUL-2002; 2002US-00187741.
XX
PR 18-SEP-1997; 97US-0059263P.
PR 18-SEP-1997; 97US-0059266P.
PR 17-OCT-1997; 97US-0062250P.
PR 21-OCT-1997; 97US-0063486P.
PR 24-OCT-1997; 97US-0063120P.
PR 24-OCT-1997; 97US-0063121P.
PR 28-OCT-1997; 97US-0063540P.
PR 28-OCT-1997; 97US-0063541P.
PR 28-OCT-1997; 97US-0063544P.
PR 28-OCT-1997; 97US-0063564P.
PR 29-OCT-1997; 97US-0063734P.
PR 31-OCT-1997; 97US-0063870P.
PR 31-OCT-1997; 97US-0064103P.
PR 13-NOV-1997; 97US-0065311P.
PR 24-NOV-1997; 97US-00656120P.
PR 24-NOV-1997; 97US-0065646P.
PR 24-NOV-1997; 97US-0065772P.
PR 11-DEC-1997; 97US-0069135P.
PR 12-DEC-1997; 97US-0069425P.
PR 17-DEC-1997; 97US-0069870P.
PR 18-DEC-1997; 97US-0068017P.
PR 10-MAR-1998; 98US-0077450P.
PR 11-MAR-1998; 98US-0077632P.
PR 11-MAR-1998; 98US-0077649P.
PR 20-MAR-1998; 98US-0078866P.

```


PR	20-MAR-1998;	98US-0078939P.	PR	24-JUN-1998;	98US-0090535P.
PR	27-MAR-1998;	98US-0079664P.	PR	24-JUN-1998;	98US-0090540P.
PR	27-MAR-1998;	98US-0079786P.	PR	25-JUN-1998;	98US-0090676P.
PR	31-MAR-1998;	98US-0080107P.	PR	25-JUN-1998;	98US-0090678P.
PR	31-MAR-1998;	98US-0080134P.	PR	25-JUN-1998;	98US-0090688P.
PR	01-APR-1998;	98US-0080327P.	PR	25-JUN-1998;	98US-0090690P.
PR	01-APR-1998;	98US-0080333P.	PR	25-JUN-1998;	98US-0090694P.
PR	08-APR-1998;	98US-0081049P.	PR	25-JUN-1998;	98US-0090695P.
PR	08-APR-1998;	98US-0081070P.	PR	25-JUN-1998;	98US-0090696P.
PR	09-APR-1998;	98US-0081195P.	PR	26-JUN-1998;	98US-00105413.
PR	15-APR-1998;	98US-0081838P.	PR	26-JUN-1998;	98US-0090862P.
PR	21-APR-1998;	98US-0082588P.	PR	26-JUN-1998;	98US-0090863P.
PR	21-APR-1998;	98US-0082589P.	PR	26-JUN-1998;	98US-0091010P.
PR	22-APR-1998;	98US-0082704P.	PR	01-JUL-1998;	98US-0091359P.
PR	28-APR-1998;	98US-0083222P.	PR	01-JUL-1998;	98US-0091544P.
PR	29-APR-1998;	98US-0083495P.	PR	02-JUL-1998;	98US-0091478P.
PR	29-APR-1998;	98US-0083496P.	PR	02-JUL-1998;	98US-0091486P.
PR	29-APR-1998;	98US-0083499P.	PR	02-JUL-1998;	98US-0091526P.
PR	29-APR-1998;	98US-0083559P.	PR	02-JUL-1998;	98US-0091528P.
PR	05-MAY-1998;	98US-0084366P.	PR	02-JUL-1998;	98US-0091632P.
PR	06-MAY-1998;	98US-0084414P.	PR	04-AUG-1998;	98US-0094006P.
PR	07-MAY-1998;	98US-0084639P.	PR	10-AUG-1998;	98US-0095282P.
PR	07-MAY-1998;	98US-0084640P.	PR	10-AUG-1998;	98US-0095988P.
PR	07-MAY-1998;	98US-0084643P.	PR	17-AUG-1998;	98US-0096012P.
PR	15-MAY-1998;	98US-0085579P.	PR	17-AUG-1998;	98US-0096757P.
PR	15-MAY-1998;	98US-0085580P.	PR	17-AUG-1998;	98US-0096766P.
PR	15-MAY-1998;	98US-0085582P.	PR	17-AUG-1998;	98US-0096867P.
PR	15-MAY-1998;	98US-0085700P.	PR	17-AUG-1998;	98US-0096891P.
PR	18-MAY-1998;	98US-0086023P.	PR	17-AUG-1998;	98US-0096897P.
PR	22-MAY-1998;	98US-0086392P.	PR	18-AUG-1998;	98US-0096949P.
PR	22-MAY-1998;	98US-0086486P.	PR	18-AUG-1998;	98US-0096959P.
PR	28-MAY-1998;	98US-0087098P.	PR	18-AUG-1998;	98US-0097022P.
PR	28-MAY-1998;	98US-0087208P.	PR	26-AUG-1998;	98US-0097552P.
PR	02-JUN-1998;	98US-0087609P.	PR	26-AUG-1998;	98US-0097954P.
PR	02-JUN-1998;	98US-0087759P.	PR	26-AUG-1998;	98US-0097971P.
PR	03-JUN-1998;	98US-0087827P.	PR	26-AUG-1998;	98US-0097974P.
PR	04-JUN-1998;	98US-0088028P.	PR	26-AUG-1998;	98US-0098014P.
PR	04-JUN-1998;	98US-0088029P.	PR	01-SEP-1998;	98US-0098716P.
PR	04-JUN-1998;	98US-0088032P.	PR	01-SEP-1998;	98US-0098723P.
PR	04-JUN-1998;	98US-0088033P.	PR	02-SEP-1998;	98US-0098803P.
PR	04-JUN-1998;	98US-0088326P.	PR	02-SEP-1998;	98US-0098821P.
PR	05-JUN-1998;	98US-0088167P.	PR	02-SEP-1998;	98US-0098843P.
PR	05-JUN-1998;	98US-0088202P.	PR	09-SEP-1998;	98US-0099602P.
PR	05-JUN-1998;	98US-0088212P.	PR	10-SEP-1998;	98US-0099741P.
PR	05-JUN-1998;	98US-0088217P.	PR	10-SEP-1998;	98US-0099754P.
PR	09-JUN-1998;	98US-0088555P.	PR	10-SEP-1998;	98US-0099763P.
PR	10-JUN-1998;	98US-0088722P.	PR	10-SEP-1998;	98US-0099812P.
PR	10-JUN-1998;	98US-0088738P.	PR	15-SEP-1998;	98US-0100388P.
PR	10-JUN-1998;	98US-0088740P.	PR	16-SEP-1998;	98US-0100662P.
PR	10-JUN-1998;	98US-0088811P.	PR	16-SEP-1998;	98US-0100664P.
PR	10-JUN-1998;	98US-0088824P.	PR	16-SEP-1998;	98US-0101751P.
PR	10-JUN-1998;	98US-0088825P.	PR	16-SEP-1998;	98US-0101751P.
PR	10-JUN-1998;	98US-0088826P.	PR	17-SEP-1998;	98US-0100683P.
PR	11-JUN-1998;	98US-0088861P.	PR	17-SEP-1998;	98US-0100684P.
PR	11-JUN-1998;	98US-0088863P.	PR	17-SEP-1998;	98US-0100919P.
PR	11-JUN-1998;	98US-0088876P.	PR	17-SEP-1998;	98US-0100930P.
PR	12-JUN-1998;	98US-0089030P.	PR	18-SEP-1998;	98US-0100849P.
PR	12-JUN-1998;	98US-0089105P.	PR	18-SEP-1998;	98US-0101014P.
PR	16-JUN-1998;	98US-0089512P.	PR	18-SEP-1998;	98US-0101068P.
PR	16-JUN-1998;	98US-0089514P.	PR	23-SEP-1998;	98US-0101471P.
PR	17-JUN-1998;	98US-0089538P.	PR	23-SEP-1998;	98US-0101472P.
PR	17-JUN-1998;	98US-0089598P.	PR	23-SEP-1998;	98US-0101475P.
PR	17-JUN-1998;	98US-0089653P.	PR	23-SEP-1998;	98US-0101477P.
PR	18-JUN-1998;	98US-0089908P.	PR	24-SEP-1998;	98US-0101738P.
PR	19-JUN-1998;	98US-0089952P.	PR	24-SEP-1998;	98US-0101739P.
PR	22-JUN-1998;	98US-0090246P.	PR	24-SEP-1998;	98US-0101743P.
PR	22-JUN-1998;	98US-0090252P.	PR	24-SEP-1998;	98US-0101922P.
PR	22-JUN-1998;	98US-0090254P.	PR	25-SEP-1998;	98US-0101786P.
PR	24-JUN-1998;	98US-0090429P.	PR	29-SEP-1998;	98US-0102207P.
PR	24-JUN-1998;	98US-0090435P.	PR	29-SEP-1998;	98US-0102240P.
PR	24-JUN-1998;	98US-0090444P.	PR	29-SEP-1998;	98US-0102330P.
PR	24-JUN-1998;	98US-0090461P.	PR	29-SEP-1998;	98US-0102331P.

```

PR 30-SEP-1998; 98US-0102487P.
PR 30-SEP-1998; 98US-0102570P.
PR 30-SEP-1998; 98US-0102571P.
PR 01-OCT-1998; 98US-0102684P.
PR 01-OCT-1998; 98US-0102687P.
PR 02-OCT-1998; 98US-0102965P.
PR 06-OCT-1998; 98US-0103259P.
PR 06-OCT-1998; 98US-0103449P.
PR 07-OCT-1998; 98US-00168978.

Query Match      31.0%; Score 205.5; DB 6; Length 1743;
Best Local Similarity 57.6%; Pred. No. 2.9e-06;
Matches 57; Conservative 6; Mismatches 35; Indels 1; Gaps 1;

QY      5 TAAVAPISVPAPVAAA-TAAATATATATTTTAAAPVAVAAAPAPAAAPSPNTAA 63
Db      1645 TAACACATTCATATAAATATATATGATTCATAAAAAAAAAAAAAAAAAAAAA 1704

QY      64 ATAAVSPAAAGQIPAAASVASAAVAFSAAAAAQQVA 102
Db      1705 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 1743

```

Search completed: June 16, 2004, 19:12:40
 Job time : 20.8313 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

XM protein - protein search, using sw model

Run on: June 16, 2004, 19:08:12 ; Search time 5.94377 Seconds
(without alignments)
1242.060 Million cell updates/sec

Title: US-09-332-063-4

Perfect score: 663

Sequence: 1 ESNKTAAPISVPAPVAAA.....QASAPAQTAAPSAPAVAPT 143

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents AA:*

1: /cgn2_6/ptodata/2/iaa/5A_COMB.pep:*

2: /cgn2_6/ptodata/2/iaa/5B_COMB.pep:*

3: /cgn2_6/ptodata/2/iaa/6A_COMB.pep:*

4: /cgn2_6/ptodata/2/iaa/6B_COMB.pep:*

5: /cgn2_6/ptodata/2/iaa/6C_COMB.pep:*

6: /cgn2_6/ptodata/2/iaa/6D_COMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	203.5	30.7	316	4	US-09-252-991A-32957
2	200.5	30.2	442	3	US-08-834-306-52
3	200.5	30.2	442	3	US-08-993-674A-52
4	200.5	30.2	442	4	US-09-256-976-52
5	196.5	29.9	219	2	US-08-257-309B-54
6	197	29.7	282	1	US-08-403-379A-1
7	197	29.7	282	2	US-08-929-414-1
8	197	29.7	263	2	US-08-557-309B-51
9	197	29.7	263	3	US-08-834-306-51
10	197	29.7	263	3	US-08-993-674A-51
11	197	29.7	263	4	US-09-256-976-51
12	195	29.4	180	6	5273901-7
13	195	29.4	180	6	5482703-6
14	195	29.4	805	3	US-09-103-429A-4
15	185.5	28.0	786	3	US-09-103-429A-3
16	181	27.3	399	4	US-09-252-991A-22853
17	177	26.7	2972	4	US-09-579-181-2
18	177	26.7	3118	4	US-09-579-181-1
19	175	26.4	98	2	US-07-814-220-2
20	175	26.4	98	2	US-07-812-421-2
21	168	25.3	756	4	US-09-963-137-184
22	165	24.9	304	4	US-09-252-991A-23116
23	162	24.4	206	4	US-08-529-055-54
24	162	24.4	8991	4	US-08-714-741-32
25	159.5	24.1	550	4	US-09-616-289-47
26	159	24.0	178	4	US-09-252-991A-20280
27	158.5	23.9	538	4	US-09-616-289-43

28 156.5 23.6 486 1 US-08-450-360-2 Sequence 2, Appli
29 156 23.5 750 4 US-09-165-239A-4 Sequence 4, Appli
30 156 23.1 165 4 US-09-489-039A-9067 Sequence 9067, Ap
31 152.5 23.0 618 4 US-09-252-991A-27666 Sequence 27666, A
32 150.5 22.7 195 4 US-09-252-991A-20967 Sequence 20967, A
33 149.5 22.5 183 4 US-08-529-055-50 Sequence 50, Appli
34 149 22.5 1652 4 US-09-627-650B-1 Sequence 1, Appli
35 149 22.5 1652 4 US-09-436-063C-1 Sequence 1, Appli
36 147 22.2 92 4 US-09-344-529-2 Sequence 2, Appli
37 147 22.2 193 4 US-08-529-055-49 Sequence 49, Appli
38 145.5 21.9 223 3 US-09-095-855-201 Sequence 201, App
39 145.5 21.9 223 4 US-09-205-426-201 Sequence 201, App
40 144.5 21.8 214 1 US-08-217-327-4 Sequence 4, Appli
41 144.5 21.8 229 4 US-09-252-991A-29247 Sequence 29247, A
42 143 21.6 642 4 US-09-527-345-2 Sequence 2, Appli
43 143 21.6 642 4 US-09-489-039A-12434 Sequence 12434, A
44 1079 4 US-09-489-039A-7502 Sequence 7502, Ap
45 142.5 21.5 272 4 US-09-252-991A-18313 Sequence 18313, A

ALIGNMENTS

RESULT 1

US-09-252-991A-32957

; Sequence 32957, Application US/09252991A

; Patent No. 6551795

; GENERAL INFORMATION:

; APPLICANT: Marc J. Rubenfield et al.

; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS

; FILE REFERENCE: 107196.136

; CURRENT APPLICATION NUMBER: US/09252.991A

; PRIOR FILING DATE: 1999-02-18

; PRIOR APPLICATION NUMBER: US 60/074,788

; PRIOR FILING DATE: 1998-02-18

; PRIOR APPLICATION NUMBER: US 60/094,190

; PRIOR FILING DATE: 1998-07-27

; NUMBER OF SEQ ID NOS: 33142

; SEQ ID NO 32957

; LENGTH: 316

; TYPE: PRT

; ORGANISM: Pseudomonas aeruginosa

US-09-252-991A-32957

Query Match 30.7%; Score 203.5; DB 4; Length 316;
Best Local Similarity 43.9%; Pred. No. 7.2e-09;
Matches 61; Conservative 8; Mismatches 57; Indels 13; Gaps 4;

Qy 4 KTAAVAPISVPAPVAAAATAAATATATTTTWWAAAPVAVAAAAAPAAAAAPSATAA 63

Db 174 KTAAPPAKPAKAAKAPAA---KPAKKPAAKTAAPKPAKPAKPAKPAKPAKPAK 230

Qy 64 AATAVSPRAAGTIPAAASVASAAAVAPASAAAQVAPAPAPVAPALVFPAPAAA 123

Db 231 AAXAAKPAK---PRAAKPAAATAKPAKPA---KPAKKPA---AKKPAKPA 280

Qy 124 QASAPAQTAAPSAPAVAP 142

Db 281 KPAAPAASSAPAPAAATP 299

RESULT 2

US-08-834-306-52

; Sequence 52, Application US/08834306

; Patent No. 6054135

; GENERAL INFORMATION:

; APPLICANT: Reed, Steven G.

; APPLICANT: Skeiky, Yasir A.W.

; APPLICANT: Lodes, Michael J.

; APPLICANT: Houghton, Raymond L.

; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE DETECTION AND PREVENTION OF T

; NUMBER OF SEQUENCES: 65

CORRESPONDENCE ADDRESS:
ADDRESSEE: SEED AND BERRY LLP
STREET: 6300 Columbia Center, 701 Fifth Avenue
CITY: Seattle
STATE: Washington
COUNTRY: USA
ZIP: 98104-7092
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/834,306
FILING DATE: 15-APR-1997
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: Maki, David J.
REGISTRATION NUMBER: 31,392
REFERENCE/DOCKET NUMBER: 210121.422C1
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031
INFORMATION FOR SEQ ID NO: 52:
SEQUENCE CHARACTERISTICS:
LENGTH: 442 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
US-08-834-306-52

Query Match 30.2%; Score 200.5; DB 3; Length 442;
Best Local Similarity 45.8%; Pred. No. 1.7e-08;
Matches 65; Conservative 6; Mismatches 64; Indels 7; Gaps 5;

2 SNKTAAPVAPISVPVAAATAAATATATATTTTVAAPVAVAAAAAPAAAAAPSPAT 61
298 SGKSKAKA--AAPAK-AAAPAKAAAPPAKAAAP--AKAAAPPAKAAAPPAKT 352
62 AAATA-AAVSPAAAGQIPAAASVASAAAVAPSAASAAAVQVAPAPVAPALVPVPAP 120
353 AAPAKTAAPPAKAAAPPAKAAAPPAKAAAPPAKAAAP--AKAAAP 411
121 AAQAASAPAQTOAPTSAPAVAP 142
412 PAKAAAPPAKAAAPPAKAAAP 433

RESULT 3
US-08-933-674A-52
; Sequence 52, Application US/08993674A
; Patent No. 6228372
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Lodes, Michael J.
; APPLICANT: Houghton, Raymond L.
; APPLICANT: Smith, John M.
; APPLICANT: McNeill, Patricia D.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE DETECTION AND PREVENTION OF T
; FILE REFERENCE: 210121.422C3
; CURRENT APPLICATION NUMBER: US/08/834,306
; CURRENT FILING DATE: 1999-02-24
; NUMBER OF SEQ ID NOS: 81
; SOFTWARE: Patent In Release #1.0, Version #1.30
; SEQ ID NO 52
; LENGTH: 442
; TYPE: PRT
; ORGANISM: Trypanosoma cruzi
US-08-933-674A-52

Query Match 30.2%; Score 200.5; DB 3; Length 442;
Best Local Similarity 45.8%; Pred. No. 1.7e-08;
Matches 65; Conservative 6; Mismatches 64; Indels 7; Gaps 5;

2 SNKTAAPVAPISVPVAAATAAATATATATTTTVAAPVAVAAAAAPAAAAAPSPAT 61
298 SGKSKAKA--AAPAK-AAAPAKAAAPPAKAAAP--AKAAAPPAKAAAPPAKT 352
62 AAATA-AAVSPAAAGQIPAAASVASAAAVAPSAASAAAVQVAPAPVAPALVPVPAP 120
353 AAPAKTAAPPAKAAAPPAKAAAPPAKAAAPPAKAAAP--AKAAAP 411
121 AAQAASAPAQTOAPTSAPAVAP 142
412 PAKAAAPPAKAAAPPAKAAAP 433

US-08-933-674A-52

CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/993,674A
FILING DATE: 18-DEC-1997
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: Maki, David J.
REGISTRATION NUMBER: 31,392
REFERENCE/DOCKET NUMBER: 210121.422C2
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031
INFORMATION FOR SEQ ID NO: 52:
SEQUENCE CHARACTERISTICS:
LENGTH: 442 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
US-08-993-674A-52

Query Match 30.2%; Score 200.5; DB 3; Length 442;
Best Local Similarity 45.8%; Pred. No. 1.7e-08;
Matches 65; Conservative 6; Mismatches 64; Indels 7; Gaps 5;

2 SNKTAAPVAPISVPVAAATAAATATATATTTTVAAPVAVAAAAAPAAAAAPSPAT 61
298 SGKSKAKA--AAPAK-AAAPAKAAAPPAKAAAP--AKAAAPPAKAAAPPAKT 352
62 AAATA-AAVSPAAAGQIPAAASVASAAAVAPSAASAAAVQVAPAPVAPALVPVPAP 120
353 AAPAKTAAPPAKAAAPPAKAAAPPAKAAAPPAKAAAP--AKAAAP 411
121 AAQAASAPAQTOAPTSAPAVAP 142
412 PAKAAAPPAKAAAPPAKAAAP 433

RESULT 4
US-09-256-976-52
; Sequence 52, Application US/09256976
; Patent No. 6419933
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Lodes, Michael J.
; APPLICANT: Houghton, Raymond L.
; APPLICANT: Smith, John M.
; APPLICANT: McNeill, Patricia D.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE DETECTION AND PREVENTION
; FILE REFERENCE: 210121.422C3
; CURRENT APPLICATION NUMBER: US/09/256,976
; CURRENT FILING DATE: 1999-02-24
; NUMBER OF SEQ ID NOS: 95
; SOFTWARE: Patent In Ver. 2.0
; SEQ ID NO 52
; LENGTH: 442
; TYPE: PRT
; ORGANISM: Trypanosoma cruzi
US-09-256-976-52

Query Match 30.2%; Score 200.5; DB 4; Length 442;
Best Local Similarity 45.8%; Pred. No. 1.7e-08;
Matches 65; Conservative 6; Mismatches 64; Indels 7; Gaps 5;

2 SNKTAAPVAPISVPVAAATAAATATATATTTTVAAPVAVAAAAAPAAAAAPSPAT 61
298 SGKSKAKA--AAPAK-AAAPAKAAAPPAKAAAP--AKAAAPPAKAAAPPAKT 352
62 AAATA-AAVSPAAAGQIPAAASVASAAAVAPSAASAAAVQVAPAPVAPALVPVPAP 120
353 AAPAKTAAPPAKAAAPPAKAAAPPAKAAAPPAKAAAP--AKAAAP 411
121 AAQAASAPAQTOAPTSAPAVAP 142

Db 412 PAKAAAPPAKAAAPPAKAAAP 433

RESULT 5

JS-08-557-309B-54
Sequence 54, Application US/08557309B
Patent No. 5916572
GENERAL INFORMATION:
APPLICANT: Reed, Steven G.
APPLICANT: Skeiky, Yasir A.W.
APPLICANT: Lodes, Michael J.
APPLICANT: Houghton, Raymond L.
TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE DETECTION AND PREVENTION OF T
NUMBER OF SEQUENCES: 69
CORRESPONDENCE ADDRESS:
ADDRESSEE: SEED and BERRY LLP
STREET: 6300 Columbia Center, 701 Fifth Avenue
CITY: Seattle
STATE: Washington
COUNTRY: USA
ZIP: 98104-7092

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/557,309B
FILING DATE: 14-NOV-1995
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Maki, David J.
REGISTRATION NUMBER: 31,392
REFERENCE/DOCKET NUMBER: 210121.422
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 622-4900
TELEFAX: (206) 622-6031
INFORMATION FOR SEQ ID NO: 54:
SEQUENCE CHARACTERISTICS:
LENGTH: 219 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear

JS-08-557-309B-54

Query Match 29.9%; Score 198.5; DB 2; Length 219;
Best Local Similarity 45.1%; Pred. No. 1.2e-08;
Matches 65; Conservative 4; Mismatches 68; Indels 7; Gaps 4;

2Y 4 KTAAVAPISVPAPVAAAATAATATATATATTTTAAAPVAVA-----AAAPAAAAPSP 59

Db 69 KKAAPSGKSAKAAAPAKAAAPAKAAAPAKAAAPAKAAAPAKAAAPAKAAAPAKAAAP 127

2Y 60 ATAAATA-AAVSPAAGQIPAAASVASAAAVAPSAASAAAVQVAPAPAPVAPALVPVP 118

Db 128 KTAAPKTAAPPAKAAAPAKAAAPAKAAAPAKAAAPAKAAAPAKAAAPAKAAAPAKAA 186

2Y 119 APAAQAASAPQTAQTSAPAVAP 142

Db 187 APPAKAAAPPAKAAAPPAEAPAP 210

RESULT 6

JS-08-403-379A-1
Sequence 1, Application US/08403379A
Patent No. 5756662
GENERAL INFORMATION:
APPLICANT: Reed, Steven G.

TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE DETECTION
TITLE OF INVENTION: OF T. CRUZI INFECTION
NUMBER OF SEQUENCES: 9
CORRESPONDENCE ADDRESS:

ADDRESSEE: SEED and BERRY LLP
STREET: 6300 Columbia Center, 701 Fifth Avenue
CITY: Seattle
STATE: Washington
COUNTRY: USA
ZIP: 98104-7092

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/403,379A
FILING DATE: 14-MAR-1995
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Maki, David J.
REGISTRATION NUMBER: 31,392
REFERENCE/DOCKET NUMBER: 210121.406
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 622-4900
TELEFAX: (206) 622-6031

INFORMATION FOR SEQ ID NO: 1:

SEQUENCE CHARACTERISTICS:

LENGTH: 262 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

US-08-403-379A-1

Query Match 29.7%; Score 197; DB 1; Length 262;
Best Local Similarity 44.3%; Pred. No. 1.9e-08;
Matches 66; Conservative 8; Mismatches 63; Indels 12; Gaps 5;

OY 1 ESNKTAAVAPISVPAPVAAA-----ATARAATATAATTTTAAAPVAVAASAAAAA 55

Db 110 EDAARAAAKQKAAKAAKAAAPSGKSKAAIAAPAKAAAPAKAAAPAKAAAPAKAA 168

OY 56 APSPTAAATAAAVSPAAGQIPAA--AAASVASAAAVAPSAASAAAVQVAPAPVAP 113

Db 169 APKAAAPAKAAATAPAKAAAPAKAAAPAKAAAPAKAAAPAKAAAPAKAAATAPAK 228

OY 114 LVPVPAPAAQAASAPQTAQTSAPAVAP 142

Db 229 ---ATAPAKA-ATAPAKAAAPAKAAATAP 253

RESULT 7

US-08-929-414-1
Sequence 1, Application US/08929414
Patent No. 5942403
GENERAL INFORMATION:
APPLICANT: Reed, Steven G.

APPLICANT: Houghton, Raymond

APPLICANT: Skeiky, Yasir A.W.

TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE DETECTION

TITLE OF INVENTION: OF T. CRUZI INFECTION

NUMBER OF SEQUENCES: 15

CORRESPONDENCE ADDRESS:

ADDRESSEE: SEED and BERRY LLP

STREET: 6300 Columbia Center, 701 Fifth Avenue

CITY: Seattle

STATE: Washington

COUNTRY: USA

ZIP: 98104-7092

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/929,414

FILING DATE: 15-SEP-1997

```
/ CLASSIFICATION: 530
/ ATTORNEY/AGENT INFORMATION:
/ NAME: Maki, David J.
/ REGISTRATION NUMBER: 31,392
/ REFERENCE/DOCKET NUMBER: 210121.406C1
/ TELECOMMUNICATION INFORMATION:
/ TELEPHONE: (206) 622-4900
/ TELEFAX: (206) 682-6031
/ INFORMATION FOR SEQ ID NO: 1:
/ SEQUENCE CHARACTERISTICS:
/ LENGTH: 262 amino acids
/ TYPE: amino acid
/ TOPOLOGY: linear
/ MOLECULE TYPE: protein
/ US-08-929-414-1

Query Match 29.7%; Score 197; DB 2; Length 262;
Best Local Similarity 44.3%; Pred. No. 1.9e-08;
Matches 66; Conservative 8; Mismatches 63; Indels 12; Gaps 5;

QY 1 ESNKTAAVAPISVPAPVAAA-----ATAAAITATAATTTTMTWAAAPVAVAAAAAPAAAA 55
DB 111 EDAAAAAAKQKAAAKGAAPSGKSKAKAATAPAKAAAAAPAKAAAAAP-AKAAAAAPAKAAA 169
QY 56 APSPTAAATAAAVSPAAAGQIPA--AASVASAAAVAPSAASAAAVQVAPAPVPAPA 113
DB 170 APKAAAAPAKXATAPAKAAAAAPKTAAPAKAAAAAPAKAAAAAPAKAATAAPAKAAAAAPAKA 229
QY 114 LVPVPAPAAAQASAPAOQAPTSAPAVAP 142
DB 230 ---ATAPAKA-ATAPAKAAAAAPAKAATAP 254

RESULT 9
US-08-834-306-51
/ Sequence 51, Application US/08834306
/ Patent No. 6054135
/ GENERAL INFORMATION:
/ APPLICANT: Reed, Steven G.
/ APPLICANT: Skeiky, Yasir A.W.
/ APPLICANT: Lodes, Michael J.
/ APPLICANT: Houghton, Raymond L.
/ TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE DETECTION AND PREVENTION OF T
/ NUMBER OF SEQUENCES: 65
/ CORRESPONDENCE ADDRESS:
/ ADDRESSEE: SEED and BERRY LLP
/ STREET: 6300 Columbia Center, 701 Fifth Avenue
/ CITY: Seattle
/ STATE: Washington
/ COUNTRY: USA
/ ZIP: 98104-7092
/ COMPUTER READABLE FORM:
/ MEDIUM TYPE: Floppy disk
/ COMPUTER: IBM PC compatible
/ OPERATING SYSTEM: PC-DOS/MS-DOS
/ SOFTWARE: PatentIn Release #1.0, Version #1.30
/ CURRENT APPLICATION DATA:
/ APPLICATION NUMBER: US/08/834,306
/ FILING DATE: 15-APR-1997
/ CLASSIFICATION: 424
/ ATTORNEY/AGENT INFORMATION:
/ NAME: Maki, David J.
/ REGISTRATION NUMBER: 31,392
/ REFERENCE/DOCKET NUMBER: 210121.422C1
/ TELECOMMUNICATION INFORMATION:
/ TELEPHONE: (206) 622-4900
/ TELEFAX: (206) 682-6031
/ INFORMATION FOR SEQ ID NO: 51:
/ SEQUENCE CHARACTERISTICS:
/ LENGTH: 263 amino acids
/ TYPE: amino acid
/ STRANDEDNESS:
/ TOPOLOGY: linear
/ US-08-834-306-51

Query Match 29.7%; Score 197; DB 3; Length 263;
Best Local Similarity 44.3%; Pred. No. 1.9e-08;
Matches 66; Conservative 8; Mismatches 63; Indels 12; Gaps 5;

QY 1 ESNKTAAVAPISVPAPVAAA-----ATAAAITATAATTTTMTWAAAPVAVAAAAAPAAAA 55
DB 111 EDAAAAAAKQKAAAKGAAPSGKSKAKAATAPAKAAAAAPAKAAAAAP-AKAAAAAPAKAAA 169
QY 56 APSPTAAATAAAVSPAAAGQIPA--AASVASAAAVAPSAASAAAVQVAPAPVPAPA 113
DB 170 APKAAAAPAKXATAPAKAAAAAPKTAAPAKAAAAAPAKAAAAAPAKAATAAPAKAAAAAPAKA 229

RESULT 8
US-08-557-309B-51
/ Sequence 51, Application US/08557309B
/ Patent No. 5916572
/ GENERAL INFORMATION:
/ APPLICANT: Reed, Steven G.
/ APPLICANT: Skeiky, Yasir A.W.
/ APPLICANT: Lodes, Michael J.
/ APPLICANT: Houghton, Raymond L.
/ TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE DETECTION AND PREVENTION OF T
/ NUMBER OF SEQUENCES: 69
/ CORRESPONDENCE ADDRESS:
/ ADDRESSEE: SEED and BERRY LLP
/ STREET: 6300 Columbia Center, 701 Fifth Avenue
/ CITY: Seattle
/ STATE: Washington
/ COUNTRY: USA
/ ZIP: 98104-7092
/ COMPUTER READABLE FORM:
/ MEDIUM TYPE: Floppy disk
/ COMPUTER: IBM PC compatible
/ OPERATING SYSTEM: PC-DOS/MS-DOS
/ SOFTWARE: PatentIn Release #1.0, Version #1.30
/ CURRENT APPLICATION DATA:
/ APPLICATION NUMBER: US/08/557,309B
/ FILING DATE: 14-NOV-1995
/ CLASSIFICATION: 435
/ ATTORNEY/AGENT INFORMATION:
/ NAME: Maki, David J.
/ REGISTRATION NUMBER: 31,392
/ REFERENCE/DOCKET NUMBER: 210121.422
/ TELECOMMUNICATION INFORMATION:
/ TELEPHONE: (206) 622-4900
/ TELEFAX: (206) 682-6031
/ INFORMATION FOR SEQ ID NO: 51:
/ SEQUENCE CHARACTERISTICS:
/ LENGTH: 263 amino acids
/ TYPE: amino acid
/ STRANDEDNESS:
/ TOPOLOGY: linear
```

```

QY 114 LVPVPAPAAQAASAPAQTAQTSAPAVAP 142
DB 230 ---ATAPAKA-ATAPAKAAAPAKAATAP 254

RESULT 10
US-08-993-674A-51
; Sequence 51, Application US/08993674A
; Patent No. 6228372
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Lodes, Michael J.
; APPLICANT: Houghton, Raymond L.
; APPLICANT: Smith, John M.
; APPLICANT: McNeill, Patricia D.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE DETECTION AND PREVENTION OF T
; NUMBER OF SEQUENCES: 81
; CORRESPONDENCE ADDRESS:
; ADDRESS: SEED and BERRY LLP
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/993,674A
; FILING DATE: 18-DEC-1997
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.422C2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 51:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 263 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
US-08-993-674A-51

Query Match 29.7%; Score 197; DB 3; Length 263;
Best Local Similarity 44.3%; Pred. No. 1.9e-08;
Matches 66; Conservative 8; Mismatches 63; Indels 12; Gaps 5;

QY 1 ESNKTAAVAPISVPAPVAAA-----ATAAATATAATTTTWWAAAPVAVAAAAAPAAAA 55
DB 111 EDAAAAAAKQKAAAKKAAAPSGKSKAAATAPAKAAAPAKAATAAP-AGAAAAAPAKAAA 169

QY 56 APSPTAATAATAAVSPRAAGQIPA--AASVASAAAVAPSAASAAAVQVAPAPVAPAPA 113
DB 170 APAKAAAPAKAATAAPAKAAAPAKTAAAPAKAAAPAKAATAAPAKAATAAPAKAATAAPAKA 229

QY 114 LVPVPAPAAQAASAPAQTAQTSAPAVAP 142
DB 230 ---ATAPAKA-ATAPAKAAAPAKAATAP 254

RESULT 11
US-09-256-976-51
; Sequence 51, Application US/09256976
; Patent No. 6419933
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.

```

```

; APPLICANT: Lodes, Michael J.
; APPLICANT: Houghton, Raymond L.
; APPLICANT: Smith, John M.
; APPLICANT: McNeill, Patricia D.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE DETECTION AND PREVENTION
; FILE REFERENCE: 210121.422C3
; CURRENT APPLICATION NUMBER: US/09/256,976
; NUMBER OF SEQ ID NOS: 95
; SOFTWARE: Patent in ver. 2.0
; SEQ ID NO 51
; LENGTH: 263
; TYPE: PRT
; ORGANISM: Trypanosoma cruzi
; FEATURE:
; OTHER INFORMATION: Where any Xaa is an independently selected amino
; OTHER INFORMATION: acid
US-09-256-976-51

Query Match 29.7%; Score 197; DB 4; Length 263;
Best Local Similarity 44.3%; Pred. No. 1.9e-08;
Matches 66; Conservative 8; Mismatches 63; Indels 12; Gaps 5;

QY 1 ESNKTAAVAPISVPAPVAAA-----ATAAATATAATTTTWWAAAPVAVAAAAAPAAAA 55
DB 111 EDAAAAAAKQKAAAKKAAAPSGKSKAAATAPAKAAAPAKAATAAP-AGAAAAAPAKAAA 169

QY 56 APSPTAATAATAAVSPRAAGQIPA--AASVASAAAVAPSAASAAAVQVAPAPVAPAPA 113
DB 170 APAKAAAPAKAATAAPAKAAAPAKTAAAPAKAAAPAKAATAAPAKAATAAPAKAATAAPAKA 229

QY 114 LVPVPAPAAQAASAPAQTAQTSAPAVAP 142
DB 230 ---ATAPAKA-ATAPAKAAAPAKAATAP 254

RESULT 12
5273901-7
; Patent No. 5273901
; APPLICANT: JACOBSON, JAMES W.; STRAUSBERG, ROBERT L.; WILSON,
; SUSAN D.; POPE, SHARON H.; STRAUSBERG, SUSAN L.; RUFF, MICHAEL D.;
; AUGUSTINE, PATRICIA C.; DANFORTH, HARRY D.
; TITLE OF INVENTION: GENETICALLY ENGINEERED COCCIDIOSIS
; SPOROZOITE 21.5 KB ANTIGEN, AC-6B
; NUMBER OF SEQUENCES: 11
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/581,693
; FILING DATE: 12-SEP-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 215,162
; FILING DATE: 03-JUL-1988
; APPLICATION NUMBER: 746,520
; FILING DATE: 19-JUN-1985
; APPLICATION NUMBER: 627,811
; FILING DATE: 05-JUL-1984
; SEQ ID NO: 7;
; LENGTH: 180
5273901-7

Query Match 29.4%; Score 195; DB 6; Length 180;
Best Local Similarity 43.0%; Pred. No. 1.8e-08;
Matches 65; Conservative 8; Mismatches 66; Indels 12; Gaps 4;

QY 1 ESNKTAAVAPISVPAPVAAAATAAA-ITATAATTTTWWAAAPV---AVAAAAAPAAAA 55
DB 27 EKEEERAAAPAAATAAPAAATAATAATAATAATAATAATAATAATAATAATAATAATAATAATA 86

QY 56 APSPTAATAATAAVSPRAA-----AGQIPAAASVASAAAVAPSA--AAAAVQVAPAPAPA 108
DB 87 APTAATAAAGAGAAKGAKEARTETEGAGAAABAEKAKTQAATAATTAAAAAASAAAG 146

QY 109 VPAPALVVPAPAAQAASAPAQTAQTSAPAPA 139

```

b6
147 KPGCHAAATAAKAQAQKAAKAANAATAATA 177

RESULT 13
5482709-6
Patent No. 5482709
APPLICANT: JACOBSON, JAMES W.; STRAUSSBERG, ROBERT L.; WILSON,
SUSAN D.; POPE, SHARON H.; STRAUSSBERG, SUSAN L.; RUFF, MICHAEL D.;
AUGUSTINE, PATRICIA C.; DANFORTH, HARRY D.
TITLE OF INVENTION: EIMERIA ANTIGENIC COMPOSITION WHICH
ELICITS ANTIBODIES AGAINST AVIAN COCCIDIOSIS
NUMBER OF SEQUENCES: 10
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/148,432
FILING DATE: 08-NOV-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 581,693
FILING DATE: 12-SEP-1990
APPLICATION NUMBER: 215,162
FILING DATE: 05-JUL-1989
APPLICATION NUMBER: 746,520
FILING DATE: 19-JUN-1985
APPLICATION NUMBER: 627,811
FILING DATE: 05-JUL-1984
SEQ ID NO: 6:
LENGTH: 180
5482709-6

Query Match 29.4%; Score 195; DB 6; Length 180;
Best Local Similarity 43.0%; Pred. No. 1.8e-08;
Matches 65; Conservative 8; Mismatches 66; Indels 12; Gaps 4;

2Y 1 ESKNTAAVPISVPAPVAAAATATA-TATATATTITTVVAAPV-----AVAAAAAPAAAA 55
DB 27 EXEERAAAPAATAAAPAAATATATAATAATAATPAAAAAPAAAAAATAATGA 86

56 APSPTATAATAAAVSPAA-----AGQIPAAASVASAAAVPSAA-AAAVQVPAAPAP 108
DB 87 APAATAAGAAGAAAGAEARTETEGAGAGAAEAEBKATQATATTTAAARAASAAG 146

109 VPAPALVPAPAAAAQAPACTQAPTSPA 139
DB 147 KPGCHAAATAAKAQAQKAAKAANAATAATA 177

RESULT 14
US-09-103-429A-4
Sequence 4, Application US/09103429A
Patent No. 6187558
GENERAL INFORMATION:
APPLICANT: Granados, Robert R
APPLICANT: Wang, Ping
TITLE OF INVENTION: A No. 6187558el Invertebrate Intestinal Mucin
TITLE OF INVENTION: CDNA and Related Products and Methods
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESSEE: Brown, Pinnisi & Michaels, P.C.
STREET: 118 No. 6187558th Tioga
CITY: Ithaca
STATE: NY
COUNTRY: USA
ZIP: 14850
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent-In Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/103,429A
FILING DATE: 24-JUN-1998
CLASSIFICATION: 800
ATTORNEY/AGENT INFORMATION:
NAME: Michaels, Christopher A
REGISTRATION NUMBER: 34,390
REFERENCE/DOCKET NUMBER: BTI-39
TELECOMMUNICATION INFORMATION:
TELEPHONE: (607) 256-2000
TELEFAX: (607) 256-3628
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 786 amino acids

NAME: Michaels, Christopher A
REGISTRATION NUMBER: 34,390
REFERENCE/DOCKET NUMBER: BTI-39
TELECOMMUNICATION INFORMATION:
TELEPHONE: (607) 256-2000
TELEFAX: (607) 256-3628
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 805 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Trichoplusia ni
TISSUE TYPE: peritrophic membrane
US-09-103-429A-4

Query Match 29.4%; Score 195; DB 3; Length 805;
Best Local Similarity 37.7%; Pred. No. 8.4e-08;
Matches 56; Conservative 10; Mismatches 61; Indels 38; Gaps 8

QY 5 TAAVAPISVPAPVAAAATAATATAI-----TTTTTAAAPVAVAAA-----A 49
DB 529 TAIPTPAAPTAAAPTAAAPESPTIVTPTAAPTAAPTTAVPEIPTVSAPTAAFTA 588

50 APAAA--AAP-----SPATAAATAAAVSPAAAGQIPAAASVASAAAVAPSAAA 95
DB 589 ATATPAPTAAPTTAVPEIPTVTSPPTAAPTAAAPNTTIVTPTAA--PTTAAAPNTIV 647

96 AAQVQVAP--AAPAPVPAPALVFVPAPAAQ---ASAPAQOAP---TSAPAVAP 142
DB 648 TAPPTAAPTAAAPTAAPTTIVTPPTAAPTAAAPTVAHPNTTAAAPTPTSAPTTP 702

RESULT 15
US-09-103-429A-3
Sequence 3, Application US/09103429A
Patent No. 6187558
GENERAL INFORMATION:
APPLICANT: Granados, Robert R
APPLICANT: Wang, Ping
TITLE OF INVENTION: A No. 6187558el Invertebrate Intestinal Mucin
TITLE OF INVENTION: CDNA and Related Products and Methods
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESSEE: Brown, Pinnisi & Michaels, P.C.
STREET: 118 No. 6187558th Tioga
CITY: Ithaca
STATE: NY
COUNTRY: USA
ZIP: 14850
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent-In Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/103,429A
FILING DATE: 24-JUN-1998
CLASSIFICATION: 800
ATTORNEY/AGENT INFORMATION:
NAME: Michaels, Christopher A
REGISTRATION NUMBER: 34,390
REFERENCE/DOCKET NUMBER: BTI-39
TELECOMMUNICATION INFORMATION:
TELEPHONE: (607) 256-2000
TELEFAX: (607) 256-3628
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 786 amino acids

TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Trichoplusia ni
TISSUE TYPE: peritrophic membrane
S-09-103-429A-3

Query Match 28.0%; Score 185.5; DB 3; Length 786;
Best Local Similarity 37.6%; Pred. No 4.4e-07;
Matches 59; Conservative 8; Mismatches 69; Indels 21; Gaps 6;
y 5 TAAVAPISVPAPVAAATAAAITATAATI-----TTMVAAAPVAVAAA--AAPAA 53
b 529 TAIPTPAFTAAPTAAAPTAAAPESPTTVTPPTAAPTAAFTTAVPEIPIITVSAPTAAPTA 588
y 54 AAAPSPATAAATAAASVPAAGQIPAAASVASAAAVAPSAAAAAVQVAP--AAPAPVPA 111
b 589 APTAAPTAAPTTAUPEIPTTVTSPTAA--PTTAAPAPNTTVTPPTAAPTAAAPAPNTT 646
y 112 PALVPVPAPAAQ---ASAPAQTOAP---TSAPAVAP 142
b 647 VTVPPTAAPTAAAPTVAHAPNTTAAAPVTTTSAPATTP 683

search completed: June 16, 2004, 19:13:24
ob time : 6.94377 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.
OM protein - protein search, using sw model
Run on: June 16, 2004, 19:10:47 ; Search time 13.8105 Seconds
(without alignments)
2917.113 Million cell updates/sec
Title: US-09-332-063-4
Perfect score: 663
Sequence: 1 BSNKTAAPVSPAPVAAA.....QASAPQQTQTSAPAVAPT 143
Scoring table: BIOSUM62
Gapop 10.0 , Gapext 0.5
Searched: 1158786 seqs, 281726120 residues
Total number of hits satisfying chosen parameters: 1158786
Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries
Database : Published Applications AA:
1: /cgm2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep.*
2: /cgm2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pep.*
3: /cgm2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep.*
4: /cgm2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep.*
5: /cgm2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep.*
6: /cgm2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pep.*
7: /cgm2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep.*
8: /cgm2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep.*
9: /cgm2_6/ptodata/2/pubpaa/US09A_PUBCOMB.pep.*
10: /cgm2_6/ptodata/2/pubpaa/US09B_PUBCOMB.pep.*
11: /cgm2_6/ptodata/2/pubpaa/US09C_PUBCOMB.pep.*
12: /cgm2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep.*
13: /cgm2_6/ptodata/2/pubpaa/US10A_PUBCOMB.pep.*
14: /cgm2_6/ptodata/2/pubpaa/US10B_PUBCOMB.pep.*
15: /cgm2_6/ptodata/2/pubpaa/US10C_PUBCOMB.pep.*
16: /cgm2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep.*
17: /cgm2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep.*
18: /cgm2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	249.5	37.6	4640	14	US-10-184-644-75
2	249.5	37.6	4640	14	US-10-184-634-75
3	249	37.6	2846	14	US-10-184-644-169
4	249	37.6	2846	14	US-10-184-634-169
5	249	37.6	2846	14	US-10-063-685-37
6	247	37.3	1904	12	US-10-142-426-99
7	247	37.3	1904	14	US-10-123-155-99
8	247	37.3	1904	14	US-10-146-731-99
9	247	37.3	1904	14	US-10-140-472-99
10	247	37.3	1904	14	US-10-141-761-99
11	247	37.3	1904	14	US-10-142-885-99
12	247	37.3	1904	14	US-10-158-790-99
13	247	37.3	1904	15	US-10-137-871-99
14	247	37.3	1904	15	US-10-140-923-99
15	247	37.3	1904	15	US-10-141-756-99

16	247	37.3	1904	15	US-10-141-759-99	Sequence 99, Appl
17	247	37.3	1904	15	US-10-140-805-99	Sequence 99, Appl
18	247	37.3	1904	15	US-10-140-864-99	Sequence 99, Appl
19	247	37.3	2773	14	US-10-184-644-149	Sequence 149, App
20	247	37.3	2773	14	US-10-184-634-149	Sequence 149, App
21	247	37.3	2773	14	US-10-063-685-33	Sequence 33, Appl
22	247	36.5	1985	14	US-10-184-644-27	Sequence 27, Appl
23	242	36.5	1985	14	US-10-184-634-27	Sequence 27, Appl
24	240.5	36.3	3501	12	US-10-142-426-37	Sequence 37, Appl
25	240.5	36.3	3501	14	US-10-123-155-37	Sequence 37, Appl
26	240.5	36.3	3501	14	US-10-146-731-37	Sequence 37, Appl
27	240.5	36.3	3501	14	US-10-140-472-37	Sequence 37, Appl
28	240.5	36.3	3501	14	US-10-141-761-37	Sequence 37, Appl
29	240.5	36.3	3501	14	US-10-142-885-37	Sequence 37, Appl
30	240.5	36.3	3501	14	US-10-158-790-37	Sequence 37, Appl
31	240.5	36.3	3501	15	US-10-137-871-37	Sequence 37, Appl
32	240.5	36.3	3501	15	US-10-140-923-37	Sequence 37, Appl
33	240.5	36.3	3501	15	US-10-141-756-37	Sequence 37, Appl
34	240.5	36.3	3501	15	US-10-141-759-37	Sequence 37, Appl
35	240.5	36.3	3501	15	US-10-140-805-37	Sequence 37, Appl
36	240.5	36.3	3501	15	US-10-140-864-37	Sequence 37, Appl
37	239.5	36.1	2379	14	US-10-184-644-483	Sequence 483, App
38	239.5	36.1	2379	14	US-10-184-634-483	Sequence 483, App
39	239	36.0	1570	14	US-10-184-644-335	Sequence 335, App
40	239	36.0	1570	14	US-10-184-634-335	Sequence 335, App
41	236.5	35.7	1771	14	US-10-184-644-17	Sequence 17, Appl
42	236.5	35.7	1771	14	US-10-184-634-17	Sequence 17, Appl
43	231.5	34.9	496	12	US-10-142-426-533	Sequence 533, App
44	231.5	34.9	496	14	US-10-123-155-533	Sequence 533, App
45	231.5	34.9	496	14	US-10-146-731-533	Sequence 533, App

ALIGNMENTS

RESULT 1
US-10-184-644-75
; Sequence 75, Application US/10184644
; Publication NO. US2003004930A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Chen, Jian
; APPLICANT: Desnoyers, Luc
; APPLICANT: Goddard, Audrey
; APPLICANT: Goddard, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Pan, James
; APPLICANT: Smith, Victoria
; APPLICANT: Watanabe, Colin K.
; APPLICANT: Wood, William I.
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P3430R1C227
; CURRENT APPLICATION NUMBER: US/10/184,644
; CURRENT FILING DATE: 2002-06-28
; Prior Application removed - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 612
; SEQ ID NO 75
; LENGTH: 4640
; TYPE: DNA
; ORGANISM: Homo Sapien
US-10-184-644-75

Query Match 37.6%; Score 249.5; DB 14; Length 4640;
Best Local Similarity 51.4%; Pred. No. 2.8e-08;
Matches 74; Conservative 7; Mismatches 56; Indels 7; Gaps 2;
QY 5 TAAVAPISVPAPVAAAATAA---ITA---TAATITTTTAAAPVAVAAAPAAAP 57
Db 4493 TAAATGAGCTTAAATAATAAGGCATATTGTCATGTTTAAATAAAAAAAAAAAAAA 4552
QY 58 SPATAATAAAVSPAAGQIPAAAGVAAAASVAAAAPVAAAAPVAAAAPVAPVAPALPV 117

```

Db      4553 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 4613
QY      118 PAPAAAOASAPAQTOAPTSAPAVA 141
Db      4613 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 4636

RESULT 2
US-10-184-634-75
; Sequence 75, Application US/10184634
; Publication No. US2003006984A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Chen, Jian
; APPLICANT: Desnoyers, Luc
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Pan, James
; APPLICANT: Smith, Victoria
; APPLICANT: Watanabe, Colin K.
; APPLICANT: Wood, William I.
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P3430R1C217
; CURRENT APPLICATION NUMBER: US/10/184,634
; PRIOR FILING DATE: 2002-06-28
; Prior Application removed - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 612
; SEQ ID NO 75
; LENGTH: 4640
; TYPE: DNA
; ORGANISM: Homo Sapien
JS-10-184-634-75

Query Match          37.6%; Score 249.5; DB 14; Length 4640;
Best Local Similarity 51.4%; Pred. No. 2.8e-08;
Matches 74; Conservative 7; Mismatches 56; Indels 7; Gaps 2;

QY      5 TRAVAPISVPAPVAAAATAAA---ITR---TAAATTTTWAAAFAVVAAAAAPAARAP 57
Db      4493 TAAATGAGCTCTTAAAAATAAAGCATATTGTCATGTTTAAAAAAAATAAAAAAAAAA 4552
QY      58 SPATAAAATARAASPAAGQTPAASAVASAAAAPSAAAAAVQVAAPAAPVPAPALVPE 117
Db      4553 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 4612

QY      118 PAPAAAOASAPAQTOAPTSAPAVA 141
Db      4613 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 4636

RESULT 3
US-10-184-644-169
; Sequence 169, Application US/10184644
; Publication No. US20030044930A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Chen, Jian
; APPLICANT: Desnoyers, Luc
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Pan, James
; APPLICANT: Smith, Victoria
; APPLICANT: Watanabe, Colin K.
; APPLICANT: Wood, William I.
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P3430R1C227

```



```

; LENGTH: 1904
; TYPE: DNA
; ORGANISM: Homo Sapien
US-10-137-871-99

Query Match      37.3%; Score 247; DB 15; Length 1904;
Best Local Similarity 53.5%; Pred. No. 1.8e-08;
Matches 68; Conservative 7; Mismatches 52; Indels 0; Gaps 0;

QY 15 APVAAATATAATATAATTTTWWAAAPVAVAAAAAPAAAAAPSPATAATAAVSPAAA 74
DB 1770 AACAAATAAAGTGAATCCCAAAAAAATAAAAAAATAAAAAAATAAAAAA 1829

QY 75 GQTPAAASVAAASAAVAPSAAAAAVQVAPAPAPVPAPALVPVPAPAAQASAPACTOAP 134
DB 1830 AAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAA 1889

QY 135 TSAPAVA 141
DB 1890 AAAAAA 1896

RESULT 14
US-10-140-923-99
; Sequence 99, Application US/10140923
; Publication No. US20030207355A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Beresini, Maureen
; APPLICANT: Deforge, Laura
; APPLICANT: Desnoyers, Luc
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Gerritsen, Mary E.
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Sherwood, Steven
; APPLICANT: Smith, Victoria
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tamas, Daniel
; APPLICANT: Watanabe, Colin K
; APPLICANT: Wood, William
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P3330R1C200
; CURRENT APPLICATION NUMBER: US/10140923
; CURRENT FILING DATE: 2002-05-07
; Prior Application removed - See Palm or File Wrapper
; NUMBER OF SEQ ID NOS: 550
; SEQ ID NO 99
; LENGTH: 1904
; TYPE: DNA
; ORGANISM: Homo Sapien
US-10-140-923-99

Query Match      37.3%; Score 247; DB 15; Length 1904;
Best Local Similarity 53.5%; Pred. No. 1.8e-08;
Matches 68; Conservative 7; Mismatches 52; Indels 0; Gaps 0;

QY 15 APVAAATATAATATAATTTTWWAAAPVAVAAAAAPAAAAAPSPATAATAAVSPAAA 74
DB 1770 AACAAATAAAGTGAATCCCAAAAAAATAAAAAAATAAAAAAATAAAAAA 1829

QY 75 GQTPAAASVAAASAAVAPSAAAAAVQVAPAPAPVPAPALVPVPAPAAQASAPACTOAP 134
DB 1830 AAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAA 1889

QY 135 TSAPAVA 141
DB 1890 AAAAAA 1896
```

```

Query Match      37.3%; Score 247; DB 15; Length 1904;
Best Local Similarity 53.5%; Pred. No. 1.8e-08;
Matches 68; Conservative 7; Mismatches 52; Indels 0; Gaps 0;

QY 15 APVAAATATAATATAATTTTWWAAAPVAVAAAAAPAAAAAPSPATAATAAVSPAAA 74
DB 1770 AACAAATAAAGTGAATCCCAAAAAAATAAAAAAATAAAAAAATAAAAAA 1829

QY 75 GQTPAAASVAAASAAVAPSAAAAAVQVAPAPAPVPAPALVPVPAPAAQASAPACTOAP 134
DB 1830 AAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAA 1889

QY 135 TSAPAVA 141
DB 1890 AAAAAA 1896
```

```

RESULT 15
US-10-141-756-99
; Sequence 99, Application US/10141756
; Publication No. US20030207359A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Beresini, Maureen
; APPLICANT: Deforge, Laura
; APPLICANT: Desnoyers, Luc
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Gerritsen, Mary E.
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Sherwood, Steven
; APPLICANT: Smith, Victoria
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tamas, Daniel
; APPLICANT: Watanabe, Colin K
; APPLICANT: Wood, William
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P3330R1C200
; CURRENT APPLICATION NUMBER: US/10141756
; CURRENT FILING DATE: 2002-05-08
; Prior Application removed - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 550
; SEQ ID NO 99
; LENGTH: 1904
; TYPE: DNA
; ORGANISM: Homo Sapien
US-10-141-756-99

Query Match      37.3%; Score 247; DB 15; Length 1904;
Best Local Similarity 53.5%; Pred. No. 1.8e-08;
Matches 68; Conservative 7; Mismatches 52; Indels 0; Gaps 0;

QY 15 APVAAATATAATATAATTTTWWAAAPVAVAAAAAPAAAAAPSPATAATAAVSPAAA 74
DB 1770 AACAAATAAAGTGAATCCCAAAAAAATAAAAAAATAAAAAAATAAAAAA 1829

QY 75 GQTPAAASVAAASAAVAPSAAAAAVQVAPAPAPVPAPALVPVPAPAAQASAPACTOAP 134
DB 1830 AAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAA 1889

QY 135 TSAPAVA 141
DB 1890 AAAAAA 1896

Search completed: June 16, 2004, 19:17:10
Job time : 14.8105 secs
```


Query Match 33.1%; Score 219.5; DB 2; Length 1299;
 Best Local Similarity 34.7%; Pred. No. 3e-05;
 Matches 59; Conservative 25; Mismatches 47; Indels 39; Gaps 5;

QY 5 TAAVAPISVPAPVAAATA-----AAITATAATTITTTWAAAPVAVAAAAAPAAAAAP 57
 DB 328 TSAPVASTLAPVLASTAPVPSGLAPVSASVSASVSPASTSAAAITSSAPASAPAP 387
 QY 58 SP-----ATAATAAVSPAAAGQIPAAASVAVASAAVAPSA 95
 DB 388 TPLASVSTPASTVILASASIPILASALASTSAPTPAASPPAAPVI--TAPTIPASAP 445
 QY 96 AAQVQVAA---APAPVPAPALVPVAPAAACASAPAQTOQTAPAPAP 142
 DB 446 TASVPLAPASAPAPAPTPVSAEPAP-----PAPAQTOQTAKHPAQN 490

RESULT 3
 S29309
 hypothetical protein 4 (phaC2 3' region) - Pseudomonas aeruginosa (fragment)
 C:Species: Pseudomonas aeruginosa
 C>Date: 10-Mar-1994 #sequence_revision 10-Mar-1994 #text_change 07-May-1999
 C:Accession: S29309
 R:Timin, A.; Steinbuechel, A.
 A:Title: Cloning and molecular analysis of the poly(3-hydroxyalkanoic acid) gene locus
 A:Reference number: S29309; MUID:93011120; PMID:1396693
 A:Accession: S29309
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-217 <TIM>
 A:Cross-references: EXBL:X66592

Query Match 33.0%; Score 218.5; DB 2; Length 217;
 Best Local Similarity 45.0%; Pred. No. 8.3e-06;
 Matches 63; Conservative 8; Mismatches 62; Indels 7; Gaps 3;

QY 4 KTAAPVAPISVPAPVAAATAATAITATAITTTWAAAPVAVAAAAAPAAAAAPSPATAA 63
 DB 67 KTAAPVAPISVPAPVAAATAATAITATAITTTWAAAPVAVAAAAAPAAAAAPSPATAA 123
 QY 64 ATAAVSPAAAGQIPAAASVAVASAAVAPSA--PAPVPAPALVPVAPAA 122
 DB 124 AAKAAKPAK---PAAAKPAKPAKPAKPAKPAKPAKPAKPAKPAKPAKPAKPAK 180
 QY 123 ACASAPNOTQAPTAPAPAP 142
 DB 181 AKPAAPAASSAPAPAPATP 200

RESULT 4
 G83013
 polyhydroxyalkanoate synthase protein Phap PA5060 [imported] - Pseudomonas aeruginosa
 C:Species: Pseudomonas aeruginosa
 C>Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 15-Jun-2001
 C:Accession: G83013
 R:Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warren, P.; Hickey, M.J.; Bhandman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; Lim, N.; Lory, S.; Olson, M.V.
 A:Title: Complete genome sequence of Pseudomonas aeruginosa PAO1, an opportunistic pathogen
 A:Reference number: A82950; MUID:20437337; PMID:10984043
 A:Accession: G83013
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-309 <STO>
 A:Cross-references: GB:AE004919; GB:AE004091; NID:g9951346; PIDN:AG08445.1; GSPDB:GN001
 A:Experimental source: strain PAO1
 A:Genetics:
 A:Gene: phap; PA5060

Query Match 33.0%; Score 218.5; DB 2; Length 309;

Best Local Similarity 45.0%; Pred. No. 1.1e-05;
 Matches 63; Conservative 8; Mismatches 62; Indels 7; Gaps 3;

QY 4 KTAAPVAPISVPAPVAAATAATAITATAITTTWAAAPVAVAAAAAPAAAAAPSPATAA 63
 DB 159 KTAAPVAPISVPAPVAAATAATAITATAITTTWAAAPVAVAAAAAPAAAAAPSPATAA 215
 QY 64 ATAAVSPAAAGQIPAAASVAVASAAVAPSA--PAPVPAPALVPVAPAA 122
 DB 216 AAKAAKPAK---PAAAKPAKPAKPAKPAKPAKPAKPAKPAKPAKPAKPAKPAK 272
 QY 123 ACASAPNOTQAPTAPAPAP 142
 DB 273 AKPAAPAASSAPAPAPATP 292

RESULT 5
 A47282
 calcium-binding protein calphotin - fruit fly (Drosophila melanogaster)
 C:Species: Drosophila melanogaster
 C>Date: 21-Sep-1993 #sequence_revision 18-Nov-1994 #text_change 21-Jul-2000
 C:Accession: A47282
 R:Martin, J.H.; Benzer, S.; Rudnicka, M.; Miller, C.A.
 A:Title: Calphotin: a Drosophila photoreceptor cell calcium-binding protein.
 A:Reference number: A47282; MUID:93165729; PMID:8094559
 A:Accession: A47282
 A:Status: preliminary
 A:Molecule type: nucleic acid
 A:Residues: 1-865 <MAR>
 A:Cross-references: GB:L02111; NID:G157031; PIDN:AAA28405.1; PID:G157032
 A:Experimental source: photoreceptor cells
 A:Note: sequence extracted from NCBI backbone (NCBIN:124955, NCBI:124956)
 C:Genetics:
 A:Gene: flyBase:Cpn
 A:Cross-references: FlyBase:FBgn0010218
 A:Superfamily: collagen alpha 1(I) chain; fibrillar collagen carboxyl-terminal homology
 C:Keywords: calcium binding

Query Match 31.7%; Score 210; DB 2; Length 865;
 Best Local Similarity 37.8%; Pred. No. 7e-05;
 Matches 70; Conservative 14; Mismatches 53; Indels 48; Gaps 8;

QY 6 AAAPVAPISVPAPVAAATAATAITATAITTTWAAAPVAVAAAAAPAAAAAPSPATAA 46
 DB 16 APVTPSAVAPVAVAAAPVAVAAAPVAVAAAPVAVAAAPVAVAAAPVAVAAAPVAVAA 75
 QY 47 ---AAAAAPAAAAAPSPATAA-----ATAAVSPAAAG-----QIPAAAAVA 84
 DB 76 TPVAVAPVAVAAAPVAVAAAPVAVAAAPVAVAAAPVAVAAAPVAVAAAPVAVAAAPV 135
 QY 85 SAAAVAPSA--PAPVPAPALVPVAP--PAPVPAPALVPVAP--PAPVPAPALVPVAP 137
 DB 136 ATPPVAAAPVAVAAAPVAVAAAPVAVAAAPVAVAAAPVAVAAAPVAVAAAPVAVAA 195
 QY 138 PAVAP 142
 DB 196 PAVAP 200

RESULT 6
 T42567
 tegument protein 24 - equine herpesvirus 4 (strain NS80567)
 C:Species: equine herpesvirus 4
 A:Variety: strain NS80567
 C>Date: 11-Jan-2000 #sequence_revision 11-Jan-2000 #text_change 21-Jul-2000
 C:Accession: T42567
 R:Telford, E.A.; Watson, M.S.; Perry, J.; Cullinane, A.A.; Davison, A.J.
 A:Title: The DNA sequence of equine herpesvirus-4.
 A:Reference number: T42567; MUID:98264497; PMID:9603335
 A:Accession: T42567
 A:Status: preliminary; translated from GB/EMBL/DDEJ

regulatory protein alqp - Pseudomonas aeruginosa
C:Species: Pseudomonas aeruginosa
C:Date: 30-Nov-1990 #sequence_revision 03-Feb-1994 #text_change 15-Oct-1999
C:Accession: A36128; JQ0148
R:Deretic, V.; Kenyecani, W.M.
J. Bacteriol. 172, 5544-5554, 1990
A:Title: A prokaryotic regulatory factor with a histone H1-like carboxy-terminal domain:
A:Reference number: A36128; MUID:91008921; PMID:1698761
A:Accession: A36128
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-352 <DER>
A:Cross-references: GB:M57551; GB:M36050
A:Experimental source: isolate cystic fibrosis
A:Note: the authors translated the codon GCC for residue 311 as Ser
R:Kato, J.; Chu, L.; Kitano, K.; DeVault, J.D.; Kimbara, K.; Chakrabarty, A.M.; Misra, I
Gene 84, 31-38, 1989
A:Title: Nucleotide sequence of a regulatory region controlling alginate synthesis in P
A:Reference number: JQ0132; MUID:90108714; PMID:2514124
A:Accession: JQ0148
A:Status: translation not shown
A:Molecule type: DNA
A:Residues: 'MGATTSARHV', 1-156, 'NA', 159-173, 178-180, 'A', 182-222, 'NA', 225-241, 'V', 243-25
C:Keywords: DNA binding; transcription regulation

Query Match 28.5%; Score 189; DB 2; Length 352;
Best Local Similarity 41.1%; Fred. No. 0.00047;
Matches 58; Conservative 6; Mismatches 75; Indels 2; Gaps 2;

QY 4 KTAAGPISVSPAPVAAAATAAATA-TAATTTTVAAPVAVAAAAAPAAAAAPSATA 62
DB 177 KTAAGKPAKPAKPAKPAKPAKPAKPAKPAKPAKPAKPAKPAKPAKPAKPAK 236

QY 63 AATAAASPAAGQIPAAASVASAAVAAPSAAAAAQA-PAAPVPAPALVPVPAPA 121
DB 237 PAAPKPAKPAKPAKPAKPAKPAKPAKPAKPAKPAKPAKPAKPAKPAKPAK 296

QY 122 AAQASAPAOQOAPTSAPAVAP 142
DB 297 AKPAKPVAAKPAKPAKPAKPAK 317

RESULT 15
A45155
mucin FIM-C.1 - African clawed frog (fragment)
C:Species: Xenopus laevis (African clawed frog)
C:Date: 26-May-1994 #sequence_revision 26-May-1994 #text_change 21-Jul-2000
C:Accession: A45155
R:Hauser, F.; Hoffmann, W.
J. Biol. Chem. 267, 24620-24624, 1992
A:Title: P-domains as shuffled cysteine-rich modules in integumentary mucin C.1 (FIM-C.1
A:Reference number: A45155; MUID:93077556; PMID:1447205
A:Accession: A45155
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-662 <HAU>
A:Cross-references: GB:L02115; NID:G214147; PIDN:AAA74725.1; PID:G951460
C:Superfamily: trefoil homology
F:162-202/Domain: trefoil homology <TRP1>
F:307-347/Domain: trefoil homology <TRP2>
F:354-394/Domain: trefoil homology <TRP3>
F:526-566/Domain: trefoil homology <TRP4>
F:573-613/Domain: trefoil homology <TRP5>
F:621-661/Domain: trefoil homology <TRP6>

Query Match 28.0%; Score 185.5; DB 2; Length 662;
Best Local Similarity 40.4%; Fred. No. 0.0012;
Matches 57; Conservative 10; Mismatches 57; Indels 17; Gaps 4;

QY 19 AATAAATATATAATTTTVAAPVAVAAAA-----APAAAAAPSATAATAAAS 70
DB 1 APTTAAVAATGKOTTAAGSAAAEKTAAGVSAAPPTAAVATGEGDATTAAATAAET 60

QY 71 PAAAGQIP--AAASVASAAVAAPSAAAAAQA-PAAP-----VPAPALVPVPAPA 122
DB 61 TTAAGEAPTTTTAPATTAAAGKAPTTAAATAPTTA-AAGAPTTATGKADATAAAPVPTTAA 119
QY 123 AQASAPAOQOAPTSAPAVAPT 143
DB 120 SNAPTAAAAATHSTAAAAAPT 140

Search completed: June 16, 2004, 19:15:40
Job time : 6.76895 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

% protein - protein search, using sw model

run on: June 16, 2004, 19:12:42 ; Search time 3.49633 seconds

(without alignments)

2129.669 Million cell updates/sec

Title: US-09-332-063-4

Perfect score: 663

Sequence: 1 ESNKTAAPVAPVAAA.....QASAPAQTAQTSAPAVAPT 143

Scoring table: BLOSUM62

Gapop 10.0, Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_42.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

result No.	Score	Query Match	Length	DB ID	Description
1	210	31.7	865	1 CPN DROME	Q02910 drosophila
2	207	31.2	439	1 XP2 XENLA	P17437 xenopus lae
3	192	29.0	316	1 CDNC HUMAN	P49018 homo sapien
4	185.5	28.0	662	1 MUC1 XENLA	Q05049 xenopus lae
5	183.5	27.7	299	1 RNP2 DROME	P50897 drosophila
6	183.5	27.7	790	1 ALP2 NOTCO	P24856 notothenia
7	181	27.3	352	1 ALGP_PSEAE	P15276 pseudomonas
8	176.5	26.6	797	1 VGLX_HSVB	P28968 equine herp
9	175	26.4	546	1 NUP1 DROME	Q9vdv3 drosophila
10	168	25.3	1300	1 SAL3 HUMAN	Q9bxa9 homo sapien
11	167.5	25.3	901	1 A180 MOUSE	Q61548 mus musculu
12	166.5	25.1	915	1 A180 RAT	Q05140 rattus norv
13	166	25.0	163	1 CU38_LOCM1	P04375 locusta mig
14	166	25.0	386	1 HXAD_MOUSE	Q62424 mus musculu
15	165.5	25.0	1046	1 IF2 STRAW	Q82K53 streptomyce
16	165.5	25.0	1083	1 T2D3 HUMAN	Q00268 homo sapien
17	164	24.7	388	1 HKAD_HUMAN	P13171 homo sapien
18	164	24.7	676	1 ICPO_HSVBJ	P29128 bovine herp
19	164	24.7	907	1 A180_HUMAN	Q60641 homo sapien
20	159	24.0	477	1 HXZ_HUMAN	P56270 homo sapien
21	159.5	23.9	416	1 ODO2_ALCEU	P52993 alcaligenes
22	158	23.8	518	1 TPW4 DROME	P49458 drosophila
23	157	23.7	1033	1 IF2 STRO	Q8c1g8 streptomyce
24	156.5	23.6	699	1 VGLG_HSV2H	P13290 herpes simp
25	154	23.2	721	1 YK82_MYCTU	Q10690 mycobacteri
26	153	23.1	553	1 ODP2_ALCEU	Q59098 alcaligenes
27	153	23.1	676	1 ICPO_HSVBK	P29836 bovine herp
28	152.5	23.0	1004	1 IF2 CORGL	Q8np40 corynebacte
29	152	22.9	331	1 MAZ MESAU	P56670 mesocricetu
30	152	22.9	477	1 MAZ_MOUSE	P56671 mus musculu
31	151	22.8	361	1 IF35_MOUSE	Q9dch4 mus musculu
32	149.5	22.5	1446	1 IE18_PRVKA	P33479 pseudorabie
33	148.5	22.4	3421	1 TEGU_HSVB	P28955 equine herp

ALIGNMENTS

RESULT 1

CPN_DROME	CPN_DROME	STANDARD;	PRT;	865 AA.
34	148	22.3	3828	1 TRX DROVI
35	147.5	22.2	672	1 HMW3 MYCPN
36	147.5	22.2	3149	1 TEGU_EBV
37	146.5	22.1	324	1 OSA_DROSI
38	146.5	22.1	466	1 HOBI_SCHPO
39	146	22.0	1783	1 RAA3_CHLRE
40	145	21.9	157	1 CU63_LOCM1
41	145	21.9	959	1 PTK1_RALSO
42	144.5	21.8	1367	1 AMYH_YEAST
43	144	21.7	1115	1 NCAL_MOUSE
44	143.5	21.6	1461	1 IE18_PRVIF
45	143.5	21.6	2716	1 OSA_DROME

OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Canton-S;
RX MEDLINE=93165729; PubMed=8094559;
RT Martin J.H., Benzer S., Rudnicka M., Miller C.A.;
RL "Calphotin: a Drosophila photoreceptor cell calcium-binding protein.";
RN Proc. Natl. Acad. Sci. U.S.A. 90:1531-1535(1993).
RP [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Canton-S;
RX MEDLINE=93165730; PubMed=8434015;
RA Ballinger D.G., Xue N., Harshman K.D.;
RT "A Drosophila photoreceptor cell-specific protein, calphotin, binds
RT calcium and contains a leucine zipper."
RL Proc. Natl. Acad. Sci. U.S.A. 90:1536-1540(1993).
CC -!- FUNCTION: Might function as a calcium-sequestering "sponge" to
CC regulate the amount of free cytoplasmic calcium. It binds 0.3 mole
CC of Ca(2+) per mole of protein.
CC -!- SUBUNIT: Homodimer (Probable).
CC -!- SUBCELLULAR LOCATION: Cytoplasmic; hypodense compartment.
CC -!- TISSUE SPECIFICITY: Soma and axons of photoreceptor cells of
CC compound eyes and ocelli.
CC -!- DEVELOPMENTAL STAGE: Expressed early in photoreceptor cell
CC development.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
CC or send an email to license@sib-sib.ch).
CC
CC EMBL; L02111; AAA28405.1; -;
CC EMBL; L05080; AAA28420.1; -;
CC PIR; A47282; A47282.
CC PIR; A47283; A47283.
CC Flybase; FBgn0010218; Cpn.
CC GO; GO:0005509; F:calcium ion binding; IDA.
CC Calcium-binding.
CC CONFLICT 36 36 A -> AVAPAWA (IN REF. 2).
CC CONFLICT 43 43 I -> T (IN REF. 2).
CC

Q24742 drosophila
Q50360 mycoplasma
P03186 epstein-bar
Q26632 drosophila
O74352 schizosacch
Q9fec4 chlamydomon
P45584 locusta mig
Q8xrh0 ralstonia s
P08640 saccharomyc
P13595 mus musculu
P11675 pseudorabie
Q8in94 drosophila

```

1  I -> V (IN REF. 2).
2  T -> A (IN REF. 2).
3  P -> PP (IN REF. 2).
4  VO -> AP (IN REF. 2).
5  I -> V (IN REF. 2).
6  S -> T (IN REF. 2).
7  A -> E (IN REF. 2).
8  I -> T (IN REF. 2).
9  V -> L (IN REF. 2).
10 D -> E (IN REF. 2).
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949
950
951
952
953
954
955
956
957
958
959
960
961
962
963
964
965
966
967
968
969
970
971
972
973
974
975
976
977
978
979
980
981
982
983
984
985
986
987
988
989
990
991
992
993
994
995
996
997
998
999
1000
1001
1002
1003
1004
1005
1006
1007
1008
1009
1010
1011
1012
1
```


Glover D.M.;
"From sequence to chromosome: the tip of the X chromosome of D.
melanogaster";
Science 287:2220-2222(2000).
[3]
SEQUENCE FROM N.A.
STRAIN=Berkley;
MEDLINE=20196006; Pubmed=10731132;
Adams M.D., Celisner S.E., Holt R.A., Evans C.A., Gocayne J.D.,
Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galie R.F.,
George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
Sutton G.G., Wortman J.R., Vandeil M.D., Zhang Q., Chen L.X.,
Brandon R.C., Rogers Y.-H.C., Blaize R.G., Champe M., Pfeiffer B.D.,
Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., McKlos G.L.G.,
Abril J.P., Agbayani A., An H.-U., Andrews-Pfannkuch C., Baldwin D.,
Bailow R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
Beeson K.Y., Benson P.V., Bereman B.P., Bhargava D., Bolshakov S.,
Borkova D., Botchan M.R., Bouck J.J., Brokstein P., Brotter P.,
Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A.I., Chandra I.,
Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
Dodson K., Douc L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
Durbin K.J., Evangelista C.C., Ferraz N.S., Ferreira S., Fleischmann W.,
Foster C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glaeser K.,
Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
Harris N.L., Harvey D.A., Heiman T.J., Hernandez J.R., Houck J.,
Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwan C.,
Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
Kimmel B.S., Kodira C.D., Kraft C., Kravitz S., Kul D., Lai Z.,
Lasko P., Lei Y., Leventosh A.A., Li J.H., Li Z., Liang Y., Lin X.,
Liu X., Mattel B., McIntosh T.C., McLeod M.P., McPherson D.,
Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleib J.M.,
Palazzo M., Pittman G.S., Pan S.S., Pollard J., Puri V., Reese M.G.,
Reiner K., Remington C., Saunders R.D.C., Scheeler F., Shen H.,
Shue B.C., Siden-Klamos I., Simpson M., Skupski M.P., Smith T.,
Spier E.C., Spradling A.C., Stapleton M., Strong R., Sun E.,
Swirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissbach J.,
Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
Ye J., Yeh R.-P., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
"The genome sequence of *Drosophila melanogaster*";
Science 287:2185-2195(2000).
-!- SIMILARITY: Belongs to the L22E family of ribosomal proteins.

This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL Outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>)
or send an email to license@isb-sib.ch.

EMBL; U42587; ABAB17433.1; -
EMBL; AL132792; CAB60023.1; -
EMBL; AE003418; AAF45546.1; -
FlyBase; FBgn0015288; RpL22.
InterPro; IPR002671; Ribosomal_L22e.
Pfam; PF01776; Ribosomal_L22e; 1.
ProDom; PD007306; Ribosomal_L22e; 1.
Ribosomal protein.
DOMAIN 24 31 POLY-ALA.
DOMAIN 46 50 POLY-ALA.
DOMAIN 65 70 POLY-ALA.
DOMAIN 93 98 POLY-ALA.
DOMAIN 103 112 POLY-ALA.
DOMAIN 136 152 POLY-ALA.
DOMAIN 185 188 POLY-LYS.
DOMAIN 292 299 ASP/GLU-RICH (HIGHLY ACIDIC).
SEQUENCE 299 AA; 30610 MW; 46A99005610E4EB0 CRC64;

```

>T PROPEP <1 5
>T CHAIN 6 790
>T PEPTIDE 6 19 ANTIPEEZE GLYCOPEPTIDE POLYPROTEIN.
>T PEPTIDE 23 36 AFGP8.
>T PEPTIDE 40 53 AFGP8.
>T PEPTIDE 57 70 AFGP8.
>T PEPTIDE 74 87 AFGP8.
>T PEPTIDE 91 104 AFGP8.
>T PEPTIDE 108 121 AFGP8.
>T PEPTIDE 125 138 AFGP8.
>T PEPTIDE 142 155 AFGP8.
>T PEPTIDE 159 172 AFGP8.
>T PEPTIDE 176 189 AFGP8.
>T PEPTIDE 193 206 AFGP8.
>T PEPTIDE 210 223 AFGP8.
>T PEPTIDE 227 240 AFGP8.
>T PEPTIDE 244 257 AFGP8.
>T PEPTIDE 261 274 AFGP8.
>T PEPTIDE 278 291 AFGP8.
>T PEPTIDE 295 308 AFGP8.
>T PEPTIDE 312 325 AFGP8.
>T PEPTIDE 329 342 AFGP8.
>T PEPTIDE 346 359 AFGP8.
>T PEPTIDE 363 376 AFGP8.
>T PEPTIDE 380 393 AFGP8.
>T PEPTIDE 397 410 AFGP8.
>T PEPTIDE 414 427 AFGP8.
>T PEPTIDE 431 444 AFGP8.
>T PEPTIDE 448 461 AFGP8.
>T PEPTIDE 465 478 AFGP8.
>T PEPTIDE 482 495 AFGP8.
>T PEPTIDE 499 512 AFGP8.
>T PEPTIDE 516 529 AFGP8.
>T PEPTIDE 533 546 AFGP8.
>T PEPTIDE 550 563 AFGP8.
>T PEPTIDE 567 580 AFGP8.
>T PEPTIDE 584 597 AFGP8.
>T PEPTIDE 601 614 AFGP8.
>T PEPTIDE 618 631 AFGP8.
>T PEPTIDE 635 648 AFGP8.
>T PEPTIDE 652 665 AFGP8.
>T PEPTIDE 669 682 AFGP8.
>T PEPTIDE 686 699 AFGP8.
>T PEPTIDE 703 716 AFGP8.
>T PEPTIDE 720 736 AFGP8.
>T PEPTIDE 740 756 AFGP8.
>T PEPTIDE 760 773 AFGP8.
>T PEPTIDE 777 790 AFGP8-LIKE.
>T PEPTIDE 790 AA; 71265 MW; 4C7CCADC48PE902 CRC64;
>T SEQUENCE 790 AA; 71265 MW; 4C7CCADC48PE902 CRC64;

Query Match 27.7%; Score 183.5; DB 1; Length 790;
Best Local Similarity 40.0%; Pred. No. 0.0017;
Matches 62; Conservative 10; Mismatches 66; Indels 17; Gaps 4;

QY 5 TAAVAPISVPVAAATAA---ITATAATITTTWVAAPVA--VAAAAAPAAAAAPS 58
DB 184 TPATAALNFATAATPATPATPALIPATATAATPATPALNFATAATPATATACNF 243
QY 59 PATATAATAAVSPA-----AAGQIPAAASVAAAAVAPAAAAAQQVAPAPVP 109
DB 244 AATAATPATATPALIPATATAATPATATACNFATAATPATATPATATPATATATAT 303
QY 110 PAPALVPVPDAA--AQASDAQTCATSPAPAP 142
DB 304 PATAALNFATAATPATATPATATPATATPATATPATATPATATPATATPATAT 338

```

```

16-OCT-2001 (Rel. 40, Last annotation update)
Transcriptional regulatory protein algp (Alginate regulatory protein
algr3).
GN ALGP OR ALGR3 OR PA5253.
Pseudomonas aeruginosa.
Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
Pseudomonadaceae; Pseudomonas.
NCBI_TaxID=287;
(1)
SEQUENCE FROM N.A.
STRAIN=8822;
MEDLINE=90108714; PubMed=2514124;
Kato J., Chu L., Kitano K., Devault J.D., Kimbara K.,
Chakrabarty A.M., Misra T.K.;
"Nucleotide sequence of a regulatory region controlling alginate
synthesis in Pseudomonas aeruginosa: characterization of the algr2
gene.";
Gene 84:31-38(1989).
(2)
SEQUENCE FROM N.A.
STRAIN=8882;
MEDLINE=90222135; PubMed=2109318;
Kato J., Misra T.K., Chakrabarty A.M.;
"Algr3, a protein resembling eukaryotic histone H1, regulates
alginate synthesis in Pseudomonas aeruginosa.";
Proc. Natl. Acad. Sci. U.S.A. 87:2887-2891(1990).
(3)
SEQUENCE FROM N.A.
STRAIN=8830;
MEDLINE=91008921; PubMed=1698761;
Deretic V., Konyecsi W.M.;
"A prokaryotic regulatory factor with a histone H1-like
carboxy-terminal domain: clonal variation of repeats within algp, a
gene involved in regulation of mucoidy in Pseudomonas aeruginosa.";
J. Bacteriol. 172:5544-5554(1990).
(4)
SEQUENCE FROM N.A.
STRAIN=PAO / PAO2003;
MEDLINE=90236911; PubMed=2110144;
Konyecsi W.M., Deretic V.;
"DNA sequence and expression analysis of algp and alqO, components of
the multigene system transcriptionally regulating mucoidy in
Pseudomonas aeruginosa: algp contains multiple direct repeats.";
J. Bacteriol. 172:2511-2520(1990).
(5)
SEQUENCE FROM N.A.
STRAIN=ATCC 15692 / PAO1;
MEDLINE=20437337; PubMed=10984043;
Stover C.K., Pham X.-Q.T., Erwin A.L., Mizoguchi S.D., Warrenner P.,
Hickey M.J., Brinkman P.S.L., Hufnagle W.O., Kowalik D.J., Lagrou M.,
Garber R.L., Goltry L., Tolentino E., Westbrock-Wadman S., Yuan Y.,
Brody L.L., Coulter S.N., Folger K.R., Kas A., Larbig K., Lim R.M.,
Smith K.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.T.,
Reizer J., Sailer M.H., Hancock R.E.W., Lory S., Olson M.V.;
"Complete genome sequence of Pseudomonas aeruginosa PAO1, an
opportunistic pathogen.";
Nature 406:959-964(2000).
-!- FUNCTION: THE PROMOTER FOR A CRITICAL ALGINATE BIOSYNTHETIC
GENE, ALGP, ENCODING GDP-MANNOSE DEHYDROGENASE, IS ACTIVATED ONLY
UNDER CONDITIONS REMINISCENT OF THE CYSTIC FIBROSIS LUNG (I.E.,
UNDER HIGH OSMOLARITY), AND AT LEAST TWO REGULATORY GENES, ALGP
AND ALQO, HAVE BEEN IMPLICATED IN THIS ACTIVATION PROCESS.
-!- DOMAIN: THE CARBOXY-TERMINAL BINDS TO DNA. IT IS UNKNOWN WHETHER
BINDING IS SPECIFIC OR NON-SPECIFIC.
-!- SIMILARITY: TO EUKARYOTIC HISTONES H1.
-----
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL Outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/)
or send an email to license@isb-sib.ch.

```

```

RESULT 7
ALGP_PSEAE
ID ALGP_PSEAE STANDARD; PRT; 352 AA.
AC P15276; Q8HTU1;
DT 01-APR-1990 (Rel. 14, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)

```

DR	EMBL; M86664; AB02506.1; -.
DR	PIR; H36802; VGBEXI.
DR	FinnProt; IPRO07110; Ig-like.
KW	Glycoprotein; Transmembrane; Signal.
FT	SIGNAL 1 22 GLYCOPROTEIN X. POTENTIAL.
FT	CHAIN 23 797 SER/THR-RICH.
FT	DOMAIN 23 465 POTENTIAL.
FT	TRANSMEM 766 790 N-LINKED (GLCNAC...) (POTENTIAL).
FT	CARBOHYD 590 590
SQ	SEQUENCE 797 AA; 80342 MW; 50C9E9211F5B5B2 CRC64;
Query Match	26.6%; Score 176.5; DB 1; Length 797;
Best Local Similarity	30.9%; Pred. No. 0.0041;
Matches 51; Conservative 19; Mismatches 72; Indels 23; Gaps 4;	
OY	2 SNKTAAVAPISVPAPVA-----AATAAAATATATIT--TTMWAAPVAAAAAAPAA 53 :: :
Dd	163 TSTTTTTATTTPVTASTTDDTTTAAATTAATTAATTAATTAATTAATTAATTAATTAATTA 222 :: :
OY	54 AAPSPTAAATAAAVSPPA---AGQIPAAASVASAAAAPSAASAAAQAQPAPAAPPV 110 :: :
Dd	223 ATTSATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 282 :: :
OY	111 APALVP-----VPAAPAAQAPAPQTAPTSAPAVAPT 143 :: :
Dd	283 ATSATPTSTSAAATSTPTPTSAASTAESTERTPTTDIT 327 :: :
RESULT 9	
ID NUPL DROME STANDARD; PRT; 546 AA.	
AC Q9VDV3; O8SYN7;	
DT 10-OCT-2003 (Rel. 42, Created)	
DT 10-OCT-2003 (Rel. 42, Last sequence update)	
DT 10-OCT-2003 (Rel. 42, Last annotation update)	
DE Probable nucleoporin Nup58.	
NUP58 OR C67360.	
OS Drosohila melanogaster (Fruit fly).	
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;	
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;	
OC Ephyrdoidea; Drosohilidae; Drosohila.	
OX NCBI_taxid=7227;	
RN [1]	
SEQUENCE FROM N.A.	
RC STRAIN=Berkelley;	
RP MEDLINE=196006; PubMed=10731132;	
RX Adams M.D., Celliker S.E., Holt R.A., Evans C.A., Gocayne J.D., RA Ananatzides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galie R.P., RA George K.A., Lewis S.B., Richards S., Ashburner M., Henderson S.N., RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X., RA Ranken R.C., Rogers Y.-H.C., Blazej R.G., Champo M., Pfeiffer B.D., RA Wan K.H., Doyle C., Baxter E.G., Heit G., Nelson C.R., Miklos G.L.O., RA April J.F., Agbayani A., An H.-J., Andrews-Pfannkuch C., Baldwin D., RA Ballwey R.M., Baay A., Bavendale J., Bayraktaroglu L., Beasley E.M., RA Beeson K.Y., Benos P.V., Berman B.P., Bhondari D., Bolshakov S., RA Borikova D., Botchan M.R., Bouck J.J., Brockstein P., Brotter P., RA Burris K.C., Busam D.A., Butier H., Cadieu E., Center A., Chandra I., RA Cherry J.M., Cavley S., Dahlke C., Davenport L.B., Davies P., RA de Pabloes B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M., RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P., RA Durbin K.J., Evangelista C.C., Ferrez C., Ferreira S., Fleischmann W., RA Foster C., Gabriellan A.E., Garg N.S., Gelbart W.M., Glasser K., RA Glodex A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M., RA Harris N.L., Harvey D.A., Helman T.Z., Hernandez J.R., Houck J., RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C., RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A., RA Kimmel B.E., Kodira C.K., Kraft C., Kravitz S., Kulip D., Lai Z., RA Lascko P., Lei Y., Levitsky A.A., Li J.H., Li Z., Liang Y., Lin X., RA Liu X., Mattel B., McIntosh T.C., McLeod M.P., MCPerson D., RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A., RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Neilson D., RA Neilson D.R., Neilson K.A., Nixon K., Nusskern D.R., Pacleb J.M., RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,	

```

DB      83  GATTAAPAGTAAATPAFCIPAAATSAFCAPAAATPAFGAAATPAFGAPAAATPAFGAPAA 142
QY      106  PAPVPAPALVWV-PAPAAQAQS---APAQ---TQAP-----TSAPAVAP 142
DB      143  TSAFGAPAAATTAFCGAPASTQAFCGAPAPAVGTVAPTFSPATPATPTTAP 194

RESULT 10
ID      SAL3 HUMAN
ID      SAL3 HUMAN      STANDARD;      PRT; 1300 AA.
DT      Q9BXA9; OSUGH1;
DT      28-FEB-2003 (Rel. 41, Created)
DT      28-FEB-2003 (Rel. 41, Last sequence update)
DT      10-OCT-2003 (Rel. 42, Last annotation update)
DE      Sal-like protein 3 (Zinc finger protein SALL3) (hSALL3).
OS      Homo sapiens (Human).
OC      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC      Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX      NCBI_TaxID=9606;
RX      SEQUENCE FROM N.A. (ISOFORM 3).
RP      Gooding R., Angelicheva D., Blechschmidt K., Swoboda K., Molnar M.,
RA      Tournev I., Kalydyjeva L.;
RT      "Exclusion of HSA1L3 and refinement of the region for the CCDFN
RT      gene.";
RL      Submitted (FEB-2001) to the EMBL/GenBank/DBSJ databases.
RL      [2]
RP      SEQUENCE OF 29-1300 FROM N.A. AND ALTERNATIVE SPLICING.
RX      MEDLINE=20079154; Pubmed=10610715;
RA      Kohlhaase J., Hausmann S., Stojmenovic G., Dixkens C., Bink K.,
RA      Schulz-Schaeffer W., Altmann M., Engel W.;
RT      "SALL3, a new member of the human spalt-like gene family, maps to
RT      18q23.";
RL      Genomics 62:216-222(1999).
RL      [3]
RN      VARIANT LEU-593, AND REVISIONS TO 787; 797-802; 808; 1138 AND 1141.
RA      Kohlhaase J.;
RL      Submitted (JUL-2002) to Swiss-Prot.
CC      !- FUNCTION: Probable transcription factor.
CC      !- SUBCELLULAR LOCATION: Nuclear (probable).
CC      !- ALTERNATIVE PRODUCTS:
CC      Event-Alternative splicing; Named isoforms=4;
CC      Name=3;
CC      IsoId=Q9BXA9-1; Sequence=Displayed;
CC      Name=1;
CC      IsoId=Q9BXA9-2; Sequence=VSP_006833;
CC      Note=Lacks two zinc finger domains. Major isoform with isoform
CC      2;
CC      Name=2;
CC      IsoId=Q9BXA9-3; Sequence=VSP_006832, VSP_006833;
CC      Note=Lacks two zinc finger domains. Major isoform with isoform
CC      1;
CC      Name=4;
CC      IsoId=Q9BXA9-4; Sequence=VSP_006832;
CC      !- TISSUE SPECIFICITY: Widely expressed in adult with highest levels
CC      in heart. Expressed in fetal brain (in neurons of hippocampus,
CC      cortex, mediodorsal and ventrolateral thalamic nuclei, putamen,
CC      cerebellum and brainstem).
CC      !- DEVELOPMENTAL STAGES: In fetal brain of the 24th gestational week.
CC      !- SIMILARITY: BELONGS TO THE SALL FAMILY OF C2H2-TYPE ZINC-FINGER
CC      PROTEINS.
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announcement/
CC      or send an email to license@isb-sib.ch).
CC      EMBL; AF347021; AAK18311.1;
CC      DR      EMBL; AF007421; CAB65124.1; ALT SEQ.

```

```

DR HSP; P07248; 1ARE.
DR Genew; HGNC:10527; SALL3.
DR MIM; 605079; -.
DR InterPro; IPR007887; Znf_C2H2.
DR Pfam; PF00096; zf_C2H2_10.
DR SMART; SM00355; Znf_C2H2_10.
DR PROSITE; PS00028; ZINC_FINGER_C2H2_2; 8.
DR PROSITE; PS00157; ZINC_FINGER_C2H2_2; 8.
KW Transcription regulation; DNA-binding; Nuclear protein; Zinc-finger;
KW Metal-binding; Repeat; Alternative splicing; Polymorphism.
FT ZN_FING 420 442 C2H2-TYPE.
FT ZN_FING 448 470 C2H2-TYPE.
FT ZN_FING 470 701 C2H2-TYPE.
FT ZN_FING 701 729 C2H2-TYPE.
FT ZN_FING 729 761 C2H2-TYPE.
FT ZN_FING 761 999 C2H2-TYPE.
FT ZN_FING 977 1027 C2H2-TYPE.
FT ZN_FING 1005 1027 C2H2-TYPE.
FT ZN_FING 1113 1135 C2H2-TYPE.
FT ZN_FING 1141 1163 C2H2-TYPE.
FT DOMAIN 198 201 POLY-ALA.
FT DOMAIN 201 216 POLY-GLN.
FT DOMAIN 213 216 POLY-SER.
FT DOMAIN 397 902 Missing (in isoform 2 and isoform 4).
FT VARSPLIC 1 133 /FTID=VSP_006832.
FT VARSPLIC 973 1044 Missing (in isoform 1 and isoform 2).
FT VARIANT 593 593 V -> L. /FTID=VSP_006833.
FT CONFLICT 235 235 R -> C (IN REF. 2). /FTID=VAR_014132.
SQ SEQUENCE 1300 AA; 195371 MW; 82CF3BDCBD59150 CRC64;
Query Match 25.3%; Score 168; DB 1; Length 1300;
Best Local Similarity 28.8%; Pred. No. 0.016;
Matches 59; Conservative 19; Mismatches 51; Indels 76; Gaps 7;
QY 10 PISVPAPVAATAAATATATATTTTVAAPVA-----VAAAAP----- 51
DB 148 PPAAPAPTPAYGAPSTNTVLEALLSTRVAVAFSQGARAAGSGAGGVAAAVPLILE 207
QY 52 -----AAAAAPSATAATAAAVSPAA 73
DB 208 QLMALQQQIHLQLEIRISQVALMQRPPLSLPAAAPSAGPAPSQPLGLAALPLS 267
QY 74 AGQIPAAASVASAAVAPSAASAAA-----AVQVAPAPAPVPAPALVPV 117
DB 268 AG-AFAAA-IAGSGPAAPEGAQPLSRPSGASTGPGPAEPSAPAAAPSAPAPA-APA 324
QY 118 PAPAQAASAPATCAPTSAPAVAP 142
DB 325 PAP-APQAASQPOSASTPPALAP 348
RESULT 11
AL80_MOUSE
ID AL80_MOUSE STANDARD; PRT; 901 AA.
AC Q61548; Q61547; PRT; 901 AA.
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Clathrin coat assembly protein AP180 (Clathrin coat associated protein
DE AP180) (91 kDa synaptosomal-associated protein) (Phosphoprotein P1-
DE 20)
DE SNAP1.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORMS LONG AND SHORT).
RX MEDLINE=92300439; PubMed=1607933;
RA Zhou S., Sousa R., Tannery N.H., Lafer E.M.;
RT "Characterization of a novel synapse-specific protein. II. cDNA
RT cloning and sequence analysis of the F1-20 protein.";
```

```

RL J. Neurosci. 12:2144-2155(1992).
CC -!- FUNCTION: Adaptins are components of the adaptor complexes which
CC link clathrin to receptors in coated vesicles. Clathrin-associated
CC protein complexes are believed to interact with the cytoplasmic
CC tails of membrane proteins, leading to their selection and
CC concentration. Binding of AP180 to clathrin triskelia induces
CC their assembly into 60-70 nm coats.
CC -!- SUBCELLULAR LOCATION: Component of the coat surrounding the
CC cytoplasmic face of coated vesicles in the plasma membrane.
CC -!- ALTERNATIVE PRODUCTS:
CC Event=Alternative splicing; Named isoforms=2;
CC Name=Long;
CC IsoId=Q61548-1; Sequence=Displayed;
CC Name=Short;
CC IsoId=Q61548-2; Sequence=VSP_000172;
CC -!- TISSUE SPECIFICITY: Brain. Associated with the synapses.
CC -!- DEVELOPMENTAL STAGE: Developmentally regulated in a pattern
CC coincident with active synaptogenesis and synaptic maturation.
CC -!- DOMAIN: Possesses a three domain structure: the N-terminal 300
CC residues harbor a clathrin binding site, an acidic middle domain
CC 450 residues, interrupted by an Ala-rich segment, and the C-
CC terminal domain (166 residues).
CC -!- PTM: Phosphorylated.
CC -!- SIMILARITY: Contains 1 epsin N-terminal homology (ENTH) domain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.ebi.ac.uk/announcements
CC or send an email to license@ebi.ac.uk).
CC -----
CC EMBL; M83985; AAA37587.1; -.
CC EMBL; M83985; AAA37586.1; -.
CC PIR; A44825; A44825.
CC MGD; MGI:109132; Snap91.
CC InterPro; IPR01026; ENTH.
CC InterPro; IPR008943; PI_bind_N.
CC Pfam; PF01417; ENTH; 1.
CC SMART; SM00273; ENTH; 1.
CC PROSITE; PS50342; ENTH; 1.
CC Coated pits; Alternative splicing; Phosphorylation.
KW DOMAIN 14 145 ENTH.
FT DOMAIN 410 413 POLY-THR.
FT DOMAIN 535 539 POLY-ALA.
FT DOMAIN 547 550 POLY-ALA.
FT DOMAIN 659 664 POLY-SER.
FT DOMAIN 704 710 POLY-SER.
FT VARSPLIC 715 719 Missing (in isoform Short).
FT /FTID=VSP_000172.
SQ SEQUENCE 901 AA; 91851 MW; 24A98F8ACE8DB8B1 CRC64;
Query Match 25.3%; Score 167.5; DB 1; Length 901;
Best Local Similarity 37.0%; Pred. No. 0.013;
Matches 60; Conservative 9; Mismatches 64; Indels 29; Gaps 6;
QY 9 APISVP-APVAAATAAATAITATATTTTVAAPVAV-----AAAAAPAAAAPS----- 58
DB 396 APISDPFAPSPPTTTTTPASASATTTAVTTTEVDLFGDAFAASPGAPAAAPSGAT 455
QY 59 -PATAATAAASVSPAAGQIPAAASVASAAAV-----APSAASAAAVQVAPAAPVPA 111
DB 456 APATAPVAAALD-ACSGNDPFPAPSESAEAPELDLFAMKPPETSAVVTTTASTAPPV 514
QY 112 PALVPVPAP-----AAAQASAPATCAPTSAPAVAP 142
DB 515 PATADSPAPTAVATAATTTAAAAATTTTAAAAATTTTAAAAATTAAP 556
RESULT 12
AL80_RAT
ID AL80_RAT STANDARD; PRT; 915 AA.
```

AC Q05140;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 15-MAR-2004 (Rel. 43, Last annotation update)
 DE Clathrin coat assembly protein AP180 (Clathrin coat associated protein)
 DE AP180 (91 kDa synaptosomal-associated protein).
 GN SNAP21.
 OS Rattus norvegicus (Rat).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 OX NCBI_TaxID=10116;
 RN [1]
 RP SEQUENCE FROM N.A. (ISOFORMS LONG AND SHORT).
 RC TISSUE=Brain;
 RX MEDLINE=93178442; PubMed=8440257;
 RA Morris S.A., Schroeder S., Plessmann U., Weber K., Ungewickell E.;
 RT "Clathrin assembly protein AP180: primary structure, domain
 organization and identification of a clathrin binding site.";
 RL EMBO J. 12:667-675 (1993).
 CC -!- FUNCTION: Adaptins are components of the adaptor complexes which
 link clathrin to receptors in coated vesicles. Clathrin-associated
 protein complexes are believed to interact with the cytoplasmic
 tails of membrane proteins, leading to their selection and
 concentration. Binding of AP180 to clathrin triskelia induces
 their assembly into 60-70 nm coats.
 CC -!- SUBCELLULAR LOCATION: Component of the coat surrounding the
 cytoplasmic face of coated vesicles in the plasma membrane.
 CC -!- ALTERNATIVE PRODUCTS:
 CC Event=Alternative splicing; Named isoforms=2;
 CC Name=Long;
 CC IsoId=Q05140-1; Sequence=Displayed;
 CC Name=Short;
 CC IsoId=Q05140-2; Sequence=VSP_000173;
 CC -!- DOMAIN: Possesses a three domain structure: the N-terminal 300
 residues harbor a clathrin binding site, an acidic middle domain
 450 residues, interrupted by an Ala-rich segment, and the C-
 terminal domain (166 residues).
 CC -!- PTM: Phosphorylated (By similarity).
 CC -!- SIMILARITY: Contains 1 epsin N-terminal homology (ENTH) domain.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@sib-sib.ch).

 DR EMBL; X68877; CAA48748.1; -;
 DR EMBL; X68878; CAA48749.1; -;
 DR FIR; S36326; S36326.
 DR FIR; S36327; S36327.
 DR InterPro; IPR001026; ENTH.
 DR Pfam; PF01417; ENTH; 1.
 DR SMART; SM00273; ENTH; 1.
 DR PROSITE; PS50942; ENTH; 1.
 KW Coated pits; Alternative splicing; Phosphorylation.
 FT DOMAIN 14 145 ENTH.
 FT DOMAIN 410 413 POLY-THR.
 FT DOMAIN 535 539 POLY-ALA.
 FT DOMAIN 547 550 POLY-ALA.
 FT DOMAIN 678 683 POLY-SER.
 FT DOMAIN 723 729 POLY-SER.
 FT VARSPIC 614 632 Missing (in isoform Short).
 SQ SEQUENCE 915 AA; 93518 MW; 32EC1B38C5DF8C0 CRC64;
 /FTID=VSP_000173.
 Query Match 25.1%; Score 166.5; DB 1; Length 915;
 Best Local Similarity 37.0%; Pred. No. 0.015;
 Matches 50; Conservative 9; Mismatches 64; Indels 29; Gaps 6;
 QY 9 APISVP-APVAAAATAATATATATTTTWWAAPVAV-----AAAAAPAAAAAPS----- 58

Db 396 APISDFFAPEPSPPTTTTTPASASATTAATTEVDLEGDFAASPGEPAPSEGAT 455
 QY 59 -PATAATAAAVSPAAGQIPAAASVASAAV-----APSAASAAAVQVAPAAPVPA 111
 Db 456 APATAPVAAALD-ACSGNDPFAPSEGSAAAPBLDLFAMKPPETSAFVVVPTTASTAPPV 514
 QY 112 PALVVPAP-----AAQASAPACAOAPTSPAPAVAP 142
 Db 515 PATAPSPAPTAAATAATTTAAATTTTSSAAATTAAP 556
 RESULT 13
 CU38 LOCKI
 ID CU38 LOCKI STANDARD; PRT; 163 AA.
 AC P04375;
 DT 20-MAR-1987 (Rel. 04, Created)
 DT 20-MAR-1987 (Rel. 04, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Cuticle protein 38 (LM-38) (LM-ACP 38).
 DE Locusta migratoria (Migratory locust).
 OS Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
 OC Neoptera; Orthopteroidea; Orthoptera; Caelifera; Acridomorpha;
 OC Acridoidea; Acrididae; Oedipodinae; Locusta.
 OX NCBI_TaxID=7004;
 RN [1]
 RP SEQUENCE.
 RX MEDLINE=87075646; PubMed=3790088;
 RA Hoeirup P., Andersen S.O., Roepstorff P.;
 RT "Primary structure of a structural protein from the cuticle of the
 migratory locust, Locusta migratoria.";
 RL Biochem. J. 236:713-720 (1986).
 RN [2]
 RP SEQUENCE OF 1-30.
 RX MEDLINE=86108304; PubMed=3943519;
 RA Hoeirup P., Andersen S.O., Roepstorff P.;
 RT "Isolation, characterization, and N-terminal sequence studies of
 cuticular proteins from the migratory locust, Locusta migratoria.";
 RL Eur. J. Biochem. 154:153-159 (1986).
 CC -!- FUNCTION: Component of the cuticle of migratory locust which
 contains more than 100 different structural proteins.
 CC -!- DOMAIN: The tetrapeptide (A-A-P-[AV]) repeats found throughout the
 protein are also present in many proteins constituting the
 protective envelope of other species.
 CC -!- SIMILARITY: TO LM-8, LM-67 AND LM-70.
 DR FIR; A03327; UCLQ38.
 KW Structural protein; Cuticle; Repeat.
 FT REPEAT 7 10
 FT REPEAT 13 16 2.
 FT REPEAT 20 23 3.
 FT REPEAT 26 29 4.
 FT REPEAT 56 59 5.
 FT REPEAT 62 65 6.
 FT REPEAT 68 71 7.
 FT REPEAT 75 78 8.
 FT REPEAT 81 84 9.
 FT REPEAT 93 96 10.
 FT REPEAT 123 126 11.
 FT REPEAT 135 138 12.
 FT REPEAT 141 144 13.
 FT REPEAT 156 159 14.
 SQ SEQUENCE 163 AA; 15320 MW; FB8C99E2CA66F49 CRC64;
 Query Match 25.0%; Score 166; DB 1; Length 163;
 Best Local Similarity 42.0%; Pred. No. 0.0044;
 Matches 55; Conservative 14; Mismatches 52; Indels 10; Gaps 4;
 QY 13 VPAPVAAAATAATATATATTTTWWAAPVAVAAAAAPAAAAAPSPATAATAAVSPA 72
 Db 6 IAAPVGYAAPAVGYAAPA-----IAAAPVAVAVAPAAASVANTYRISQTAARVLAAP 58
 QY 73 ANQQIPAAASVASAAAVAPSAASAAAVQVAP-AAAPVAPALVPVAPAA-AQASAPAPQ 130

```

Db      59  AYAAPAAVAAAPGUYAATAAATAAALGARYAAAAAFV-AVAHAAPAAASVANTYRISQ 111
Qy      131  TOAPTASAPAA 141
Db      118  TARTLAAAPAA 128

RESULT 14
ID      HXAD_MOUSE
AC      Q62424;
CD      15-DEC-1998 (Rel. 37, Created)
DT      DT
DT      15-DEC-1998 (Rel. 37, Last sequence update)
DE      15-MAR-2004 (Rel. 43, Last annotation update)
DE      Homeobox protein Hox-A13 (Hox-1.10).
GN      HOXA13 OR HOX-1.10.
OS      Mus musculus (Mouse).
OC      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC      Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX      NCBI_TaxID=10090;
RN      [1]
RP      SEQUENCE FROM N.A.
MD      MEDLINE=96259555; PubMed=8673126;
RA      Mortlock D.P., Post L.C., Innis J.W.;
RT      "The molecular basis of hypodactyly (Hd): a deletion in Hoxa 13 leads
RT      to arrest of digital arch formation.";
RL      Nat. Genet. 13:284-289(1996)
CC      -!- FUNCTION: Sequence-specific transcription factor which is part of
CC      a developmental regulatory system that provides cells with
CC      specific positional identities on the anterior-posterior axis (By
CC      similarity).
CC      -!- SUBCELLULAR LOCATION: Nuclear.
CC      -!- DISEASE: Defects in HOXA13 are the cause of hypodactyly (Hd), a
CC      condition characterized by profound deficiency of digital arch
CC      structures.
CC      -!- SIMILARITY: Belongs to the Atd-B homeobox family.

-----
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (see http://www.isb-sib.ch/announcement
or send an email to license@sb-sib.ch).
-----
EMBL; U59322; AAB03322.1; -
JR      HSSP; P14653; 1B72.
JR      TRANSFAC; T03337; -
JR      MGI; 96173; Hoxa13.
JR      InterPro; IPR001356; Homeobox.
JR      Pfam; PF00046; homeobox; 1.
JR      ProDom; PD000010; Homeobox; 1.
JR      SMART; SM00389; HOX; 1.
JR      PROSITE; PS00027; HOMEBOX 1; 1.
JR      PROSITE; PS50071; HOMEBOX 2; 1.
JR      Homeobox; DNA-binding; Developmental protein; Nuclear protein;
JR      Transcription regulation.
JR      DOMAIN 38 51 POLY-ALA.
JR      DNA BIND 320 379 HOMEBOX.
JR      DOMAIN 52 57 POLY-GLY.
JR      DOMAIN 62 66 POLY-ALA.
JR      DOMAIN 73 84 POLY-ALA.
JR      DOMAIN 101 104 POLY-ALA.
JR      DOMAIN 116 133 POLY-ALA.
JR      DOMAIN 198 205 POLY-ALA.
JR      SEQUENCE 386 AA; 2B01DCC9B1951324 CRC64;

Query Match 25.0%; Score 166; DB 1; Length 386;
Best Local Similarity 40.0%; Pred.No. 0.0084;
Matches 56; Conservative 12; Mismatches 50; Indels 22; Gaps 4;

Y      18  AAAATAAAITATAATI-----TTTAAAPVAAAAAFAAAAAAPS-----PATA 62

```

```
DR Pfam: PF03144; GTP EFTU D2; 1.
DR Pfam: PF04760; IF2_N; 2.
DR PRINTS: PRO03115; ELONGATNCT.
DR PRINTS: PRO12117; PRICHXTNSN.
DR PRINTS: PRO0449; RASTRNSFRNG.
DR ProDom: PD186100; IF2; 1.
DR SMART: SM00173; RAS; 1.
DR TIGRFAMs: TIGR00487; IF-2; 1.
DR TIGRFAMs: TIGR00231; small GTP; 1.
DR PROSITE: PS01176; IF2; FALSE NEG.
KW Initiation factor; Protein biosynthesis; GTP-binding;
KW Complete proteome.
FT DOMAIN 542 694 G-DOMAIN.
FT NP_BIND 548 555 GTP (BY SIMILARITY).
FT NP_BIND 598 602 GTP (BY SIMILARITY).
FT NP_BIND 652 655 GTP (BY SIMILARITY).
SQ SEQUENCE 1046 AA; 106882 MW; 62B12E2DFA3AA595 CRC64;

Query Match 25.0%; Score 165.5; DB 1; Length 1046;
Best Local Similarity 43.3%; Pred. No. 0.019;
Matches 58; Conservative 4; Mismatches 47; Indels 25; Gaps 7;

2Y 9 APISVPAPVAAAATAAATAATATATTTTVAAPVAVAAAAAPAAAAAPSPATAAATAA 68
Db |::||| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
34 ASSTIEAVVVKLTDLQCGNGG-----KAAP-----RKAAPAKCPAQAPAPAP 82
2Y 69 VSPAAGCIPAAASVASAAAVAPSAAAAAAVOVAPAPAPVPAPALVVPAPAAAAQASAP 128
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
83 RPPAPK---PAAAE-RPAAEPPAA-----APAAPGPPGPKPAPRPAPAPAPAPAP 130
2Y 129 AQTOAPTSAPAVAP 142
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
131 EFT-APSPAP-AP 142
```

Search completed: June 16, 2004, 19:17:45
Job time : 4.49633 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

3M protein - protein search, using sw model

Run on: June 16, 2004, 19:17:48 ; Search time 51 seconds
(without alignments)

4175.976 Million cell updates/sec

Title: US-09-332-063-2

Perfect score: 675

Sequence: 1 MPRAQSSASQVPADPEA.....KTPILQGPDAEMVEVLI 675

Scoring table: OLIGO

Gapop 60.0 , Gapext 60.0

Searched: 1017041 seqs, 315518202 residues

Word size : 0

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :

```

1: sp_archea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_virus:*
16: sp_bacteriap:*
17: sp_archeap:*

```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB ID	Description
1	675	100.0	675	4 Q9HD27	Q9hd27 homo sapien
2	103	15.3	891	11 Q8VHG2	Q8vhg2 mus musculus
3	24	3.6	326	4 Q8NDNO	Q8ndno homo sapien
4	24	3.6	498	4 Q96CM5	Q96cm5 homo sapien
5	24	3.6	882	11 Q9D4H4	Q9d4h4 mus musculus
6	24	3.6	956	4 Q8IY63	Q8iy63 homo sapien
7	18	2.7	402	4 Q8TBR8	Q8tbr8 homo sapien
8	18	2.7	463	11 Q8BP84	Q8bp84 mus musculus
9	18	2.7	463	11 Q8BS08	Q8bs08 mus musculus
10	18	2.7	463	11 Q7IPE4	Q7ipe4 mus musculus
11	18	2.7	466	4 Q96F99	Q96f99 homo sapien
12	18	2.7	569	4 Q9UKB4	Q9ukb4 homo sapien
13	18	2.7	586	11 Q9QUS0	Q9qus0 mus musculus
14	18	2.7	772	11 Q8K371	Q8k371 mus musculus
15	18	2.7	859	4 Q9Y2J4	Q9y2j4 homo sapien
16	13	1.9	557	4 Q8WDX1	Q8wdx1 homo sapien

ALIGNMENTS

RESULT 1

```

Q9HD27
ID Q9HD27 PRELIMINARY; PRT; 675 AA.
AC Q9HD27; Q9UPT1;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
GN Angiomotin (Hypothetical protein KIAA1071).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21157403; PubMed=11257124;
RA Troyanovsky B., Levchenko T., Mansson G., Matvijenko O., Holmgren L.;
RT "Angiomotin. An angiotensin binding protein that regulates endothelial
RT cell migration and tube formation.";
RL J. Cell Biol. 152:1247-1254(2001).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=99397452; PubMed=10470851;
RA Kikuno R., Nagase T., Ishikawa K., Hirose M., Miyajima N.,
RA Tanaka A., Kotani H., Nomura N., Ohara O.;
RT "Prediction of the coding sequences of unidentified human genes. XIV.
RT The complete sequences of 100 new cDNA clones from brain which code
RT for large proteins in vitro.";
RL DNA Res. 6:197-205(1999).
DR EMBL; AF286598; AAG01851.1; ..
DR EMBL; AB028994; BAA83023.3; ..
KW Hypothetical protein.
SQ SEQUENCE 675 AA; 72540 MW; BCC28B74427AD481 CRC64;

```

Query Match 100.0%; Score 675; DB 4; Length 675;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 675; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

Q8ten8 homo sapien
Q60586 mastomys hi
Q8vnp9 pseudomonas
Q7w8g5 bordetella
Q7wm27 bordetella
Q9qzps rattus norv
Q7xwm7 oryza sativ
Q9a3h3 caulobacter
Q8wqj2 spodoptera
Q8afb0 pseudomonas
Q8p5b3 xanthomonas
Q7xq14 oryza sativ
Q9yh51 pseudopleur
Q8t179 methanosarc
Q8py50 methanosarc
Q8evj1 mycoplasma
Q8fsa0 corynebacte
Q96q14 homo sapien
Q7wf90 bordetella
Q7w3w0 bordetella
Q7vus8 bordetella
Q18507 tenebrio mo
Q9a8h3 caulobacter
Q8pl97 xanthomonas
Q93qt0 mycobacteri
Q93qt1 mycobacteri
Q963b8 toxoplasma
Q86g86 toxoplasma
Q8mqto aedes albop

```

QY 1 MPRAQPSASQVPVADPFAIVSRAQQMVEILSDENRNLRLQELGGCYEKVARLQKVETEI 60
 DB 1 MPRAQPSASQVPVADPFAIVSRAQQMVEILSDENRNLRLQELGGCYEKVARLQKVETEI 60
 QY 61 QVSEAYENLVKSSKRALEKAVNKLEGEIRHMDPNEDLREELTANKQIAEKEYEG 120
 DB 61 QVSEAYENLVKSSKRALEKAVNKLEGEIRHMDPNEDLREELTANKQIAEKEYEG 120
 QY 121 SEDTRKTIQSOLFPAKNKESQREKELEAEATATARSTNEDQRRHIEIRDOALSNAQAQVVKL 180
 DB 121 SEDTRKTIQSOLFPAKNKESQREKELEAEATATARSTNEDQRRHIEIRDOALSNAQAQVVKL 180
 QY 181 EEEELKKQVYVDKVEKMQQALVOLQACEKEBQLEHLRLTRELRELESRLIOQRQNCOP 240
 DB 181 EEEELKKQVYVDKVEKMQQALVOLQACEKEBQLEHLRLTRELRELESRLIOQRQNCOP 240
 QY 241 TNVSEYNAALMELLREKEERIILALEADMTKWEQKYLEENVMRHFALDAAATVAAGRDIT 300
 DB 241 TNVSEYNAALMELLREKEERIILALEADMTKWEQKYLEENVMRHFALDAAATVAAGRDIT 300
 QY 301 VISHSPNTSYDTALEARTQKEEIEILMANKRCCLDMGRITKLHAQIIEKIDAMIKVLQORS 360
 DB 301 VISHSPNTSYDTALEARTQKEEIEILMANKRCCLDMGRITKLHAQIIEKIDAMIKVLQORS 360
 QY 361 RKEPSKTEQLSCMPAKSLMSISNAGSGLSHSSTLTGSPIMEEKRDDKSWKSGSLGILLG 420
 DB 361 RKEPSKTEQLSCMPAKSLMSISNAGSGLSHSSTLTGSPIMEEKRDDKSWKSGSLGILLG 420
 QY 421 GDYRAEYVPSPTSPVPPSTPILLSHSKTSGRDCSTQTERGTESNKTAAVAPISVPAPVAA 480
 DB 421 GDYRAEYVPSPTSPVPPSTPILLSHSKTSGRDCSTQTERGTESNKTAAVAPISVPAPVAA 480
 QY 481 AATAAATATATATTTTVAAPAVAAAAAPAAAAAPSPATAAATAAASVPAAGQIPA 540
 DB 481 AATAAATATATTTTVAAPAVAAAAAPAAAAAPSPATAAATAAASVPAAGQIPA 540
 QY 541 AASVASAAAVAPSAAAAVQVAPAPVAPALVPVPAAPAAQASAPAQTAQTSAPA 600
 DB 541 AASVASAAAVAPSAAAAVQVAPAPVAPALVPVPAAPAAQASAPAQTAQTSAPA 600
 QY 601 VAPTPAPTPTPAVQAQVAPSPATGPGPHRLISPLTNCNPKDTPGVFHSNTLERKTPIQ 660
 DB 601 VAPTPAPTPTPAVQAQVAPSPATGPGPHRLISPLTNCNPKDTPGVFHSNTLERKTPIQ 660
 QY 661 ILGQEPDAEMVEYLI 675
 DB 661 ILGQEPDAEMVEYLI 675
 RESULT 2
 Q8VHG2
 ID Q8VHG2 PRELIMINARY; PRT; 891 AA.
 AC Q8VHG2;
 DT 01-WAR-2002 (TRENBLrel. 20, Created)
 DT 01-WAR-2002 (TRENBLrel. 20, Last sequence update)
 DT 01-OCT-2003 (TRENBLrel. 25, Last annotation update)
 DE Angiomotin.
 GN AMOT.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=B6CBF2; TISSUE=Placenta;
 RA Troyanovsky B., Bratt A., Holmgren L.;
 RT "Mouse angiomotin."
 RL Submitted (DEC-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AF461135; AAL73436.1; --
 XR MGD; MGI:108440; Amot.
 SQ SEQUENCE 891 AA; 95111 MW; C0544542A4CD6088 CRC64;
 Query Match 15.3%; Score 103; DB 11; Length 891;

Best Local Similarity 100.0%; Pred. No. 3e-83;
 Matches 103; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 244 SEYRAAALMELLREKEERIILALEADMTKWEQKYLEENVMRHFALDAAATVAAGRDITV 303
 DB 398 SEYRAAALMELLREKEERIILALEADMTKWEQKYLEENVMRHFALDAAATVAAGRDITV 457
 QY 304 HSPNTSYDTALEARTQKEEIEILMANKRCCLDMGRITKLHAQI 346
 DB 458 HSPNTSYDTALEARTQKEEIEILMANKRCCLDMGRITKLHAQI 500
 RESULT 3
 Q8NDNO
 ID Q8NDNO PRELIMINARY; PRT; 326 AA.
 AC Q8NDNO;
 DT 01-OCT-2002 (TRENBLrel. 22, Created)
 DT 01-OCT-2002 (TRENBLrel. 22, Last sequence update)
 DT 01-OCT-2002 (TRENBLrel. 22, Last annotation update)
 DE Hypothetical protein (Fragment).
 GN DKFZP434O2216.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Testis;
 RA Koeherer K., Beyer A., Mewes H.W., Gassenhuber J., Wiemann S.;
 RL Submitted (JUL-2002) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AL833833; CAD38693.1; --
 XW Hypothetical protein.
 FT NON TER
 SQ SEQUENCE 326 AA; 35950 MW; SE28183735B67B56 CRC64;
 Query Match 3.6%; Score 24; DB 4; Length 326;
 Best Local Similarity 100.0%; Pred. No. 9.2e-13;
 Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 256 REKEERILALEADMTKWEQKYLEE 279
 DB 41 REKEERILALEADMTKWEQKYLEE 64
 RESULT 4
 Q96CM5
 ID Q96CM5 PRELIMINARY; PRT; 498 AA.
 AC Q96CM5;
 DT 01-DEC-2001 (TRENBLrel. 19, Created)
 DT 01-DEC-2001 (TRENBLrel. 19, Last sequence update)
 DT 01-DEC-2001 (TRENBLrel. 19, Last annotation update)
 DE Similar to RIKEN cDNA 4932416D09 gene (Fragment).
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Skin;
 RA Strausberg R.;
 RL Submitted (SEP-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL; BC014126; AAL14126.1; --
 FT NON TER
 SQ SEQUENCE 498 AA; 56333 MW; AD557A70AB2EDDD4 CRC64;
 Query Match 3.6%; Score 24; DB 4; Length 498;
 Best Local Similarity 100.0%; Pred. No. 1.3e-12;
 Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 256 REKEERILALEADMTKWEQKYLEE 279
 DB 213 REKEERILALEADMTKWEQKYLEE 236

```

Best Local Similarity 100.0%; Pred. No. 2,3e-12;
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 256 REKERILALEADMTKWEQKYLEE 279
Db 671 REKERILALEADMTKWEQKYLEE 694

RESULT 7
Q8TB88 PRELIMINARY; PRT; 402 AA.
ID Q8TB88
AC Q8TB88
DT 01-JUN-2002 (TrEMBLrel. 21, Created)
DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE Hypothetical protein (Fragment).
OS Homo sapiens (Human)
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
SEQUENCE FROM N.A.
RC TISSUE=Pancreas;
RA Strausberg R.;
RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC025981; AAH25981.1; -
KW Hypothetical protein.
FT NON-TER 1
SQ SEQUENCE 402 AA; 44234 MW; 5AD4ADFOE8A83676 CRC64;

Query Match 2.7%; Score 18; DB 4; Length 402;
Best Local Similarity 100.0%; Pred. No. 2.6e-07;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 262 ILALEADMTKWEQKYLEE 279
Db 182 ILALEADMTKWEQKYLEE 199

RESULT 8
Q8BP84 PRELIMINARY; PRT; 463 AA.
ID Q8BP84
AC Q8BP84
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Leman coiled-coil protein.
GN AMOTL2.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
[1]
RN SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Body;
RX MEDLINE=22354683; PubMed=12466851;
RA The FANTOM Consortium,
RA The RIKEN Genome Exploration Research Group Phase I & II Team;
RT "Analysis of the mouse transcriptome based on functional annotation of
RT 60,770 full-length cDNAs.";
RL Nature 420:563-573 (2002).
DR EMBL; AK075335; BAC36853.1; -
DR PIR; PT0534; PT0534.
DR MGD; MGI:1929286; Amotl2.
SQ SEQUENCE 463 AA; 51733 MW; 0394E0334ACFAL7E CRC64;

Query Match 2.7%; Score 18; DB 11; Length 463;
Best Local Similarity 100.0%; Pred. No. 3e-07;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 262 ILALEADMTKWEQKYLEE 279
Db 247 ILALEADMTKWEQKYLEE 264

Best Local Similarity 100.0%; Pred. No. 2.1e-12;
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 256 REKERILALEADMTKWEQKYLEE 279
Db 596 REKERILALEADMTKWEQKYLEE 619

RESULT 6
Q8IY63 PRELIMINARY; PRT; 956 AA.
ID Q8IY63
AC Q8IY63
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Similar to angiotensin like 1.
OS Homo sapiens (Human)
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
[1]
RN SEQUENCE FROM N.A.
RC TISSUE=Muscle;
RA Strausberg R.;
RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC037539; AAH37539.1; -
DR MGD; MGI:1929286; Amotl2.
SQ SEQUENCE 956 AA; 106574 MW; 6F47AE3A13816E95 CRC64;

Query Match 3.6%; Score 24; DB 4; Length 956;
Best Local Similarity 100.0%; Pred. No. 2.1e-12;
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 256 REKERILALEADMTKWEQKYLEE 279
Db 596 REKERILALEADMTKWEQKYLEE 619

RESULT 5
Q9D4H4 PRELIMINARY; PRT; 892 AA.
ID Q9D4H4
AC Q9D4H4
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE 4932416D09Rik protein.
GN AMOTL1 OR 4932416D09RIK.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
[1]
RN SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Testis;
RX MEDLINE=21085660; PubMed=11217851;
RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,
RA Saito T., Okazaki Y., Gojohori T., Bono H., Kasukawa T., Saito R.,
RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA Fleischmann W., Gasterland T., Gissi C., King B., Kochiwa H.,
RA Kuehl P., Lewis S., Matsuo Y., Mikaido I., Pesole G., Quackenbush J.,
RA Schiraldi L.M., Straubli P., Suzuki R., Tomita M., Wagner L., Washio T.,
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA Sasaki H., Sato K., Schoenbach C., Seva T., Shibata Y., Storch K.-P.,
RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
RA Wyszewski-Boris A., Yoshida K., Hasegawa Y., Kawai H., Kohetsuki S.,
RA Hayashizaki Y.;
RA "Functional annotation of a full-length mouse cDNA collection.";
RT Nature 409:685-690 (2001).
RL EMBL; AK016526; BAB30287.1; -
DR MGD; MGI:1929273; Amotl1.
DR GO; GO:0005923; C:right junction; IDA
SQ SEQUENCE 892 AA; 98423 MW; 88AF83FFA56046 CRC64;

Query Match 3.6%; Score 24; DB 11; Length 892;
Best Local Similarity 100.0%; Pred. No. 2.1e-12;
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 256 REKERILALEADMTKWEQKYLEE 279
Db 596 REKERILALEADMTKWEQKYLEE 619

```

RESULT 9							
ID	QBSB08 PRELIMINARY; PRT; 463 AA.						
DT	QBS08;						
DT	01-MAR-2003 (TrEMBLrel. 23, Created)						
DT	01-MAR-2003 (TrEMBLrel. 23, Last sequence update)						
DT	01-OCT-2003 (TrEMBLrel. 25, Last annotation update)						
DE	Leman coiled-coil protein.						
EN	AMOTL2.						
OS	Mus musculus (Mouse).						
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;						
CC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.						
CX	NCBI_TaxID=10090;						
RN	{1}_						
RP	SEQUENCE FROM N.A.						
RC	STRAIN=C57BL/6J; TISSUE=Aorta, and Vein;						
RX	MEDLINE=22354683; PubMed=12466851;						
RA	The FANTOM Consortium,						
RT	"Analysis of the mouse transcriptome based on functional annotation of						
RL	60,770 full-length cDNAs"; Nature 420:563-573(2002);						
DR	EMBL; AK040912; BAC30740.1; -						
PIR	PTO534; PT0534..						
MGB	MG1:1929286; Amotl2.						
SEQ	SEQUENCE 463 AA; 51758 MW; 2F7BE52FBC8B1678 CRC64;						
Query Match 2.7%; Score 18; DB 11; Length 463;							
Best Local Similarity 100.0%; Pred.No. 3e-07;							
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;							
DY	262 ILALEADMTKWEQYLEE 279 						
DB	247 ILALEADMTKWEQYLEE 264 						
RESULT 10							
ID	Q7TPE4 PRELIMINARY; PRT; 463 AA.						
AC	Q7TPE4;						
DT	01-OCT-2003 (TrEMBLrel. 25, Created)						
DT	01-OCT-2003 (TrEMBLrel. 25, Last sequence update)						
DE	Angiomotin-like protein 2 variant 2.						
OS	Mus musculus (Mouse).						
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;						
CC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.						
CX	NCBI_TaxID=10090;						
RN	{1}						
RP	SEQUENCE FROM N.A.						
KP	Kessler R., Brunet J.-F., Van Meir E.G.;						
RT	"Murine homolog of the human LCCP gene, Variant 2.";						
DL	Submitted (AUG-1999) to the EMBL/GenBank/DBJ databases.						
RL	EMBL; AF175968; AADS56363.1; -						
SEQ	SEQUENCE 463 AA; 51778 MW; 1493D080B4FBBD70 CRC64;						
Query Match 2.7%; Score 18; DB 11; Length 463;							
Best Local Similarity 100.0%; Pred.No. 3e-07;							
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;							
OY	262 ILALEADMTKWEQYLEE 279 						
DB	247 ILALEADMTKWEQYLEE 264 						
RESULT 11							
ID	Q96F99 PRELIMINARY; PRT; 466 AA.						
DT	Q96F99;						
DT	01-DEC-2001 (TrEMBLrel. 19, Created)						

EMBL; AB023206; BAA76833.1; --

Search completed: June 16, 2004, 19:27:17
Job time : 54 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

MX protein - protein search, using sw model

Run on: June 16, 2004, 19:25:49 ; Search time 23 Seconds
(without alignments)
1515.110 Million cell updates/sec

Title: US-09-332-063-2

Perfect score: 675

Sequence: 1 MPRAQPSASQVPADPFA.....KTFIQLGQEPDAMVEXLI 675

Scoring table: OLIGO

Gapop 60.0 , Gapext 60.0

Searched: 389414 seqs, 51625971 residues

Word size : 0

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :

- 1: /cgn2_6/ptodata/2/iaa/5A COMB pep.*
- 2: /cgn2_6/ptodata/2/iaa/5B COMB pep.*
- 3: /cgn2_6/ptodata/2/iaa/6A COMB pep.*
- 4: /cgn2_6/ptodata/2/iaa/6B COMB pep.*
- 5: /cgn2_6/ptodata/2/iaa/6C COMB pep.*
- 6: /cgn2_6/ptodata/2/iaa/backfiles1.pcp.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	11	1.6	76	6	5273901-11
2	11	1.6	76	6	5482709-10
3	10	1.5	127	4	US-09-107-532A-6784
4	10	1.5	130	4	Sequence 6784, Ap
5	10	1.5	156	1	US-08-074-121-5
6	10	1.5	156	5	Sequence 5, Appli
7	10	1.5	170	4	US-09-252-991A-27069
8	10	1.5	180	6	Sequence 27069, A
9	10	1.5	180	6	5273901-7
10	10	1.5	308	3	US-08-413-974-2
11	10	1.5	308	3	Sequence 2, Appli
12	10	1.5	308	3	US-08-434-418-2
13	10	1.5	308	3	Sequence 2, Appli
14	10	1.5	308	3	US-08-174-729A-2
15	10	1.5	308	4	US-08-434-256-2
16	10	1.5	375	4	US-09-903-814A-2
17	9	1.3	16	2	US-09-489-039A-12434
18	9	1.3	22	6	US-08-308-494A-2
19	9	1.3	22	6	5304631-11
20	9	1.3	52	4	US-09-621-976-4003
21	9	1.3	118	3	US-08-301-162-10
22	9	1.3	118	4	US-09-461-240-10
23	9	1.3	118	4	US-09-968-927-10
24	9	1.3	130	4	US-09-540-014-6
25	9	1.3	181	4	US-09-252-991A-23483
26	9	1.3	202	4	US-09-252-991A-28326
27	9	1.3	263	4	US-09-252-991A-22230
			384	4	US-09-252-991A-24427

ALIGNMENTS

RESULT 1

5273901-11

Patent No. 5273901

; APPLICANT: JACOBSON, JAMES W.; STRAUSBERG, ROBERT L.; WILSON, SUSAN D.; POPE, SHARON H.; STRAUSBERG, SUSAN L.; RUFF, MICHAEL D.; AUGUSTINE, PATRICIA C.; DANFORTH, HARRY D.

; TITLE OF INVENTION: GENETICALLY ENGINEERED COCCIDIOSIS

; SPOOROITE 21.5 KB ANTIGEN, AC-6B

; NUMBER OF SEQUENCES: 11

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/07/581,693

; FILING DATE: 12-SEP-1990

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: 215,162

; FILING DATE: 05-JUL-1988

; APPLICATION NUMBER: 746,520

; FILING DATE: 19-JUN-1985

; APPLICATION NUMBER: 627,811

; FILING DATE: 05-JUL-1986

; SEQ ID NO:11:

; LENGTH:76

5273901-11

Query Match 1.6%; Score 11; DB 6; Length 76;

Best Local Similarity 100.0%; Pred.No. 0.017;

Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 507 AAAAAA 517

DB 19 AAAAAA 29

RESULT 2

5482709-10

Patent No. 5482709

; APPLICANT: JACOBSON, JAMES W.; STRAUSBERG, ROBERT L.; WILSON, SUSAN D.; POPE, SHARON H.; STRAUSBERG, SUSAN L.; RUFF, MICHAEL D.; AUGUSTINE, PATRICIA C.; DANFORTH, HARRY D.

; TITLE OF INVENTION: BIMERIA ANTIGENIC COMPOSITION WHICH

; ELICITS ANTIBODIES AGAINST AVIAN COCCIDIOSIS

; NUMBER OF SEQUENCES: 10

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/148,432

; FILING DATE: 08-NOV-1993

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: 581,693

; FILING DATE: 12-SEP-1990

; APPLICATION NUMBER: 215,162

; FILING DATE: 05-JUL-1989

; APPLICATION NUMBER: 746,520

Y FILING DATE: 19-JUN-1985
APPLICATION NUMBER: 627,811
FILING DATE: 05-JUL-1984
SEQ ID NO:10:
LENGTH: 76
5482709-10

Query Match
Best Local Similarity 1.6%; Score 11; DB 6; Length 76;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Y 507 AAAPAAAZAA 517
|||||||
b 19 AAAPAAAZAA 29

RESULT 3
US-09-107-532A-6784
Sequence 6784, Application US/09107532A
Patent No. 6583275
GENERAL INFORMATION:
APPLICANT: Lynn A Doucette-Stamm and David Bush
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
ENTEROCOCCUS FAECIUM FOR DIAGNOSTICS AND THERAPEUTICS
NUMBER OF SEQUENCES: 7310
CORRESPONDENCE ADDRESS:
ADDRESSEE: GENOME THERAPEUTICS CORPORATION
STREET: 100 Beaver Street
CITY: Waltham
STATE: Massachusetts
COUNTRY: USA
ZIP: 02354
COMPUTER READABLE FORM:
MEDIUM TYPE: CD-ROM ISO9660
COMPUTER: PC
OPERATING SYSTEM: <Unknown>
SOFTWARE: ASCII
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/107,532A
FILING DATE: 30-Jun-1998
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 60/085,598
FILING DATE: 14 May 1998
APPLICATION NUMBER: 60/051571
FILING DATE: July 2, 1997
ATTORNEY/AGENT INFORMATION:
NAME: Arianello, Pamela Deneke
REGISTRATION NUMBER: 40,489
REFERENCE/DOCKET NUMBER: STC-012
TELECOMMUNICATION INFORMATION:
TELEPHONE: (781)893-5007
TELEFAX: (781)893-8277
INFORMATION FOR SEQ ID NO: 6784:
SEQUENCE CHARACTERISTICS:
LENGTH: 127 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: YES
ORIGINAL SOURCE:
ORGANISM: Enterococcus faecium
FEATURE:
NAME/KEY: misc feature
LOCATION: (B) LOCATION 1...127
SEQUENCE DESCRIPTION: SEQ ID NO: 6784:
IS-09-107-532A-6784
Query Match
Best Local Similarity 1.5%; Score 10; DB 4; Length 127;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Y 500 AAAPVAVAAA 509
|||||||

Db 40 AAAPVAVAAA 49

RESULT 4
US-09-050-739-64
Sequence 64, Application US/09050739
Patent No. 6641814
GENERAL INFORMATION:
APPLICANT: ANDERSEN, Peter
APPLICANT: NIELSEN, Rikke
APPLICANT: OETTINGER, Thomas
APPLICANT: RASMUSSEN, Peter Birk
APPLICANT: ROSENKRANDS, Ida
APPLICANT: WELDKINGH, Karin
APPLICANT: FLORIO, Walter
TITLE OF INVENTION: NUCLEIC ACIDS FRAGMENTS AND POLYPEPTIDE FRAGMENTS
TITLE OF INVENTION: DERIVED FROM M. TUBERCULOSIS
FILE REFERENCE: 670001-2002.1
CURRENT APPLICATION NUMBER: US/09/050,739
CURRENT FILING DATE: 1998-03-30
EARLIER APPLICATION NUMBER: 0376/97
EARLIER FILING DATE: 1997-04-02
EARLIER APPLICATION NUMBER: 1277/97
EARLIER FILING DATE: 1997-11-10
EARLIER APPLICATION NUMBER: 60/044,624
EARLIER FILING DATE: 1997-04-18
EARLIER APPLICATION NUMBER: 60/070,488
EARLIER FILING DATE: 1998-01-05
NUMBER OF SEQ ID NOS: 173
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 64
LENGTH: 130
TYPE: PRT
ORGANISM: Mycobacterium tuberculosis
US-09-050-739-64

Query Match 1.5%; Score 10; DB 4; Length 130;
Best Local Similarity 100.0%; Pred. No. 0.24;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 500 AAAPVAVAAA 509
|||||||

Db 36 AAAPVAVAAA 45

RESULT 5
US-08-074-121-5
Sequence 5, Application US/08074121
Patent No. 5767362
GENERAL INFORMATION:
APPLICANT: Best, Elaine
APPLICANT: Knauf, Vic C.
TITLE OF INVENTION: Methods and Compositions for Modulating
Lipid Content of Plant Tissues
NUMBER OF SEQUENCES: 15
CORRESPONDENCE ADDRESS:
ADDRESSEE: Fish & Richardson
STREET: 2200 Sand Hill Road, Suite 100
CITY: Menlo Park
STATE: California
COUNTRY: US
ZIP: 94025
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/074,121
FILING DATE: 08-JUN-1993
CLASSIFICATION: 800
ATTORNEY/AGENT INFORMATION:
NAME: Rae-Venter, Barbara

```

; REGISTRATION NUMBER: 32,750
; REFERENCE/DOCKET NUMBER: 05938/043001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 854-5277
; TELEFAX: (415) 854-0875
; TELEX: 200154
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 156 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-074-121-5

Query Match 1.5%; Score 10; DB 1; Length 156;
Best Local Similarity 100.0%; Pred. No. 0.28;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

y 508 AAAAPAAAAA 517
b 60 AAAAPAAAAA 69

RESULT 6
CT-US94-06447-5
Sequence 5, Application PC/TUS9406447
GENERAL INFORMATION:
APPLICANT: Calgene, Inc.
TITLE OF INVENTION: Methods and Compositions for Modulating
TITLE OF INVENTION: Lipid Content of Plant Tissues
NUMBER OF SEQUENCES: 15
CORRESPONDENCE ADDRESSES:
ADDRESSEE: Weil, Gotshal & Manges
STREET: 2882 Sand Hill Road, Suite 280
CITY: Menlo Park
STATE: California
COUNTRY: US
ZIP: 94025
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US94/06447
FILING DATE: 06-JUN-1994
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Rae-Venter, Barbara
REGISTRATION NUMBER: 32,750
REFERENCE/DOCKET NUMBER: CGNE-097/WO
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 926-6200
TELEFAX: (415) 854-3713
TELEX:
INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 156 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; CT-US94-06447-5

Query Match 1.5%; Score 10; DB 5; Length 156;
Best Local Similarity 100.0%; Pred. No. 0.28;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

y 508 AAAAPAAAAA 517
b 60 AAAAPAAAAA 69

RESULT 7
CT-US94-06447-5
Sequence 5, Application PC/TUS9406447
GENERAL INFORMATION:
APPLICANT: Calgene, Inc.
TITLE OF INVENTION: Methods and Compositions for Modulating
TITLE OF INVENTION: Lipid Content of Plant Tissues
NUMBER OF SEQUENCES: 15
CORRESPONDENCE ADDRESSES:
ADDRESSEE: Weil, Gotshal & Manges
STREET: 2882 Sand Hill Road, Suite 280
CITY: Menlo Park
STATE: California
COUNTRY: US
ZIP: 94025
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US94/06447
FILING DATE: 06-JUN-1994
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Rae-Venter, Barbara
REGISTRATION NUMBER: 32,750
REFERENCE/DOCKET NUMBER: CGNE-097/WO
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 926-6200
TELEFAX: (415) 854-3713
TELEX:
INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 156 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; CT-US94-06447-5

Query Match 1.5%; Score 10; DB 5; Length 156;
Best Local Similarity 100.0%; Pred. No. 0.28;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

y 508 AAAAPAAAAA 517
b 60 AAAAPAAAAA 69

RESULT 8
5273901-7
Patent No. 5273901
APPLICANT: JACOBSON, JAMES W.; STRAUSBERG, ROBERT L.; WILSON,
SUSAN D.; POPE, SHARON H.; STRAUSBERG, SUSAN L.; RUFF, MICHAEL D.;
AUGUSTINE, PATRICIA C.; DANFORTH, HARRY D.
TITLE OF INVENTION: GENETICALLY ENGINEERED COCCIDIOSIS
; SPOOROITE 21.5 KB ANTIGEN, AC-6B
; NUMBER OF SEQUENCES: 11
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/581,693
; FILING DATE: 12-SEP-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 215,162
; FILING DATE: 05-JUL-1988
; APPLICATION NUMBER: 746,520
; FILING DATE: 19-JUN-1985
; APPLICATION NUMBER: 627,811
; FILING DATE: 05-JUL-1984
; SEQ ID NO: 7:
; LENGTH: 180
; 5273901-7

Query Match 1.5%; Score 10; DB 6; Length 180;
Best Local Similarity 100.0%; Pred. No. 0.33;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

y 507 AAAAPAAAAA 516
b 64 AAAAPAAAAA 73

RESULT 9
5482709-6
Patent No. 5482709
APPLICANT: JACOBSON, JAMES W.; STRAUSBERG, ROBERT L.; WILSON,
SUSAN D.; POPE, SHARON H.; STRAUSBERG, SUSAN L.; RUFF, MICHAEL D.;
AUGUSTINE, PATRICIA C.; DANFORTH, HARRY D.
TITLE OF INVENTION: EIMERIA ANTIGENIC COMPOSITION WHICH
ELICITS ANTIBODIES AGAINST AVIAN COCCIDIOSIS
; NUMBER OF SEQUENCES: 10
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/148,432
; 5482709-6

US-09-252-991A-27069
; Sequence 27069, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 27069
; LENGTH: 170
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
; US-09-252-991A-27069

Query Match 1.5%; Score 10; DB 4; Length 170;
Best Local Similarity 100.0%; Pred. No. 0.31;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

y 508 AAAAPAAAAA 517
b 74 AAAAPAAAAA 83
```



```
; FILING DATE: 08-NOV-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 581,693
; FILING DATE: 12-SEP-1990
; APPLICATION NUMBER: 215,162
; FILING DATE: 05-JUL-1989
; APPLICATION NUMBER: 746,520
; FILING DATE: 19-JUN-1985
; APPLICATION NUMBER: 627,811
; FILING DATE: 05-JUL-1984
; SEQ ID NO: 6:
; LENGTH: 180
5482709-6

Query Match
Best Local Similarity 1.5%; Score 10; DB 6; Length 180;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 507 AAAAATAAA 516
Db 64 AAAAATAAA 73

RESULT 10
JS-08-413-974-2
; Sequence 2, Application US/08413974
; Patent No. 6180368
; GENERAL INFORMATION:
; APPLICANT: Singh, Mohan Bir
; APPLICANT: Knox, Robert B.
; APPLICANT: Smith, Penelope
; APPLICANT: Avjoglou, Asil
; APPLICANT: Theerakulpisut, Piyada
; APPLICANT: Hough, Terry
; TITLE OF INVENTION: Ryegrass Pollen Allergen
; NUMBER OF SEQUENCES: 25
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Woodcock Washburn Kurtz Mackiewicz & No. 6180368ris
; STREET: 1 Liberty Place, 46th Floor
; CITY: Philadelphia
; STATE: PA
; COUNTRY: USA
; ZIP: 19103
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/413,974
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/202,861
; FILING DATE:
; APPLICATION NUMBER: US/07/746,703
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Hohnschutz, Liza D.
; REGISTRATION NUMBER: 33,712
; REFERENCE/DOCKET NUMBER: IMPH-0003
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (215) 568-3100
; TELEFAX: (215) 568-3949
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 308 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-413-974-2
```

Query Match 1.5%; Score 10; DB 3; Length 308;

```
Best Local Similarity 100.0%; Pred. No. 0.54;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 521 ATAAATAAAV 530
Db 281 ATAAATAAAV 290

RESULT 11
US-08-434-418-2
; Sequence 2, Application US/08434418
; Patent No. 6197313
; GENERAL INFORMATION:
; APPLICANT: Singh, Mohan Bir et al.
; TITLE OF INVENTION: RYEGRASS POLLEN ALLERGEN
; FILE REFERENCE: IMI-051CND2
; CURRENT APPLICATION NUMBER: US/08/434,418
; PRIOR FILING DATE: 1995-05-03
; PRIOR APPLICATION NUMBER: 08/202,861
; PRIOR FILING DATE: 1994-25-02
; NUMBER OF SEQ ID NOS: 25
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 2
; LENGTH: 308
; TYPE: PRT
; ORGANISM: Lolium perenne
US-08-434-418-2

Query Match 1.5%; Score 10; DB 3; Length 308;
Best Local Similarity 100.0%; Pred. No. 0.54;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 521 ATAAATAAAV 530
Db 281 ATAAATAAAV 290

RESULT 12
US-08-433-288-2
; Sequence 2, Application US/08433288
; Patent No. 6239269
; GENERAL INFORMATION:
; APPLICANT: Singh, Mohan Bir et al.
; TITLE OF INVENTION: RYEGRASS POLLEN ALLERGEN
; FILE REFERENCE: IMI-051CND1
; CURRENT APPLICATION NUMBER: US/08/433,288
; CURRENT FILING DATE: 1995-05-03
; PRIOR APPLICATION NUMBER: 08/413,947
; PRIOR FILING DATE: 1995-03-30
; PRIOR APPLICATION NUMBER: 08/202,861
; PRIOR FILING DATE: 1994-02-25
; PRIOR APPLICATION NUMBER: 07/746,703
; PRIOR FILING DATE: 1991-08-16
; PRIOR APPLICATION NUMBER: 07/585,086
; PRIOR FILING DATE: 1990-10-26
; NUMBER OF SEQ ID NOS: 25
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 2
; LENGTH: 308
; TYPE: PRT
; ORGANISM: Lolium perenne
US-08-433-288-2

Query Match 1.5%; Score 10; DB 3; Length 308;
Best Local Similarity 100.0%; Pred. No. 0.54;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 521 ATAAATAAAV 530
Db 281 ATAAATAAAV 290

RESULT 13
```

IS-08-174-739A-2
Sequence 2, Application US/08174739A
Patent No. 6265566
GENERAL INFORMATION:
APPLICANT: Singh, Mohan Bir
APPLICANT: Knox, Robert B.
APPLICANT: Smith, Penelope
APPLICANT: Avtioglu, Asil
APPLICANT: Theerakulpisut, Piyada
APPLICANT: Hough, Terryn
TITLE OF INVENTION: Ryegrass Pollen Allergen
NUMBER OF SEQUENCES: 26
CORRESPONDENCE ADDRESS:
ADDRESSEE: Lahive & Cockfield, LLP
STREET: 60 State Street, Suite 510
CITY: Boston
STATE: MA
COUNTRY: USA
ZIP: 02109
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/174,739A
FILING DATE: 29-DEC-1993
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Mandragoras, Amy E.
REGISTRATION NUMBER: 36,207
REFERENCE/DOCKET NUMBER: IMI-051DV
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 227-7400
TELEFAX: (617) 227-5941
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 308 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-174-739A-2

Query Match 1.5%; Score 10; DB 3; Length 308;
Best Local Similarity 100.0%; Pred. No. 0.54;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

y 521 ATAAATAAAV 530
|||||
b 281 ATAAATAAAV 290

RESULT 14
S-08-434-256-2
Sequence 2, Application US/08434256
Patent No. 6451324
GENERAL INFORMATION:
APPLICANT: Singh, Mohan Bir, Knox, Robert B., Smith, Penelope,
APPLICANT: Avtioglu, Asil, Theerakulpisut, Piyada, Hough, Terryn
TITLE OF INVENTION: Ryegrass Pollen Allergen
NUMBER OF SEQUENCES: 25
CORRESPONDENCE ADDRESS:
ADDRESSEE: Woodcock Washburn Kurtz Mackiewicz & No. 6451324-ris
STREET: 1 Liberty Place, 46th Floor
CITY: Philadelphia
STATE: PA
COUNTRY: USA
ZIP: 19103
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25

;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: US/08/434,256
;; FILING DATE:
;; CLASSIFICATION: 435
;; ATTORNEY/AGENT INFORMATION:
;; NAME: Hohenschutz, Liza D.
;; REGISTRATION NUMBER: 33,712
;; REFERENCE/DOCKET NUMBER: IMPE-0003
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: (215)568-3100
;; TELEFAX: (215)568-3949
;; INFORMATION FOR SEQ ID NO: 2:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 308 amino acids
;; TYPE: amino acid
;; TOPOLOGY: linear
;; MOLECULE TYPE: protein
US-08-434-256-2

Query Match 1.5%; Score 10; DB 4; Length 308;
Best Local Similarity 100.0%; Pred. No. 0.54;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 521 ATAAATAAAV 530
|||||
Db 281 ATAAATAAAV 290

RESULT 15
US-09-303-814A-2
Sequence 2, Application US/099030814A
Patent No. 6680428
GENERAL INFORMATION:
APPLICANT: Falco, Carl
APPLICANT: Pamodu, Lavo O.
APPLICANT: Orozco, Buddy
APPLICANT: Rafalski, Antoni
APPLICANT: Thorpe, Cathy
TITLE OF INVENTION: Tetrahydrofolate Metabolism Enzymes
FILE REFERENCE: BB1179 USDIV
CURRENT APPLICATION NUMBER: US/09/903,814A
CURRENT FILING DATE: 2001-07-12
PRIOR APPLICATION NUMBER: US 60/092,869
PRIOR FILING DATE: 1998-07-15
PRIOR APPLICATION NUMBER: US 09/351,703
PRIOR FILING DATE: 1999-07-12
NUMBER OF SEQ ID NOS: 22
SOFTWARE: Microsoft Office 97
SEQ ID NO 2
LENGTH: 375
TYPE: PRT
ORGANISM: Zea mays
US-09-303-814A-2

Query Match 1.5%; Score 10; DB 4; Length 375;
Best Local Similarity 100.0%; Pred. No. 0.66;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 474 VPAPVAAAAT 483
|||||
Db 247 VPAPVAAAAT 256

Search completed: June 16, 2004, 19:31:43
Job time : 24 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

WM protein - protein search, using sw model

Run on: June 16, 2004, 19:27:24 ; Search time 49 Seconds
(without alignments)
3880.921 Million cell updates/sec

title: US-09-332-063-2
effect score: 675
sequence: 1 MPRAQSSASVQVPADPFA.....KTPQILQGPDAENVLYLI 675

coring table: OLIGO
Gapop 60.0 , Gapext 60.0

searched: 1158786 seqs, 281726120 residues

ord size : 0

total number of hits satisfying chosen parameters: 1158786

minimum DB seq length: 0

maximum DB seq length: 2000000000

post-processing: Listing first 45 summaries

database : Published Applications AA.*

1: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep.*
2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pep.*
3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep.*
4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep.*
5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep.*
6: /cgn2_6/ptodata/2/pubpaa/PCRUS_PUBCOMB.pep.*
7: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep.*
8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep.*
9: /cgn2_6/ptodata/2/pubpaa/US09A_PUBCOMB.pep.*
10: /cgn2_6/ptodata/2/pubpaa/US09B_PUBCOMB.pep.*
11: /cgn2_6/ptodata/2/pubpaa/US09C_PUBCOMB.pep.*
12: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep.*
13: /cgn2_6/ptodata/2/pubpaa/US10A_PUBCOMB.pep.*
14: /cgn2_6/ptodata/2/pubpaa/US10B_PUBCOMB.pep.*
15: /cgn2_6/ptodata/2/pubpaa/US10C_PUBCOMB.pep.*
16: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep.*
17: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep.*
18: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

result No.	Score	Query Match	Length	ID	Description
1	55	8.1	208	9	US-09-864-761-36456
2	48	7.1	48	9	Sequence 36456, A
3	26	3.9	26	9	Sequence 39389, A
4	24	3.6	608	14	US-10-204-887-87
5	24	3.6	882	14	Sequence 87, Appl
6	18	2.7	467	15	US-10-298-417-2
7	18	2.7	467	15	Sequence 52, Appl
8	11	1.6	228	12	US-10-243-552-605
9	11	1.6	407	15	Sequence 605, App
10	11	1.6	443	15	Sequence 13871, A
11	11	1.6	1966	15	Sequence 10023, A
12	10	1.5	120	12	US-10-282-122A-60991
13	10	1.5	120	16	Sequence 3424, Ap
14	10	1.5	120	16	Sequence 6099, A
15	10	1.5	122	9	Sequence 1360, Ap
16	10	1.5	122	9	Sequence 1545, Ap
17	10	1.5	122	9	Sequence 13247, A

Sequence 13588, A
Sequence 57450, A
Sequence 6, Appli
Sequence 20, Appl
Sequence 1441, Ap
Sequence 1863, Ap
Sequence 1737, Ap
Sequence 1825, Ap
Sequence 2228, Ap
Sequence 1314, Ap
Sequence 1365, Ap
Sequence 1531, Ap
Sequence 1546, Ap
Sequence 53900, A
Sequence 7215, Ap
Sequence 14026, A
Sequence 1270, Ap
Sequence 4044, Ap
Sequence 7, Appli
Sequence 8, Appli
Sequence 1051, Ap
Sequence 64, Appl
Sequence 64, Appl
Sequence 61922, A
Sequence 62378, A
Sequence 64413, A
Sequence 64, Appl
Sequence 64, Appl
Sequence 1286, Ap
Sequence 1794, Ap

16 10 1.5 122 9 US-09-815-242-13588
17 10 1.5 122 12 US-10-282-122A-57450
18 10 1.5 122 15 US-10-386-050A-6
19 10 1.5 122 15 US-10-386-050A-20
20 10 1.5 122 16 US-10-389-566-1441
21 10 1.5 122 16 US-10-389-566-1863
22 10 1.5 123 16 US-10-389-566-1737
23 10 1.5 124 16 US-10-389-566-1825
24 10 1.5 124 16 US-10-389-566-2228
25 10 1.5 125 16 US-10-389-566-1314
26 10 1.5 125 16 US-10-389-566-1365
27 10 1.5 125 16 US-10-389-566-1531
28 10 1.5 125 16 US-10-389-566-1546
29 10 1.5 126 12 US-10-282-122A-53900
30 10 1.5 126 16 US-10-389-566-1215
31 10 1.5 127 12 US-10-282-122A-74026
32 10 1.5 127 16 US-10-389-566-1370
33 10 1.5 128 9 US-09-738-626-4044
34 10 1.5 128 13 US-10-013-379-7
35 10 1.5 128 13 US-10-013-379-8
36 10 1.5 128 16 US-10-389-566-1051
37 10 1.5 130 9 US-09-791-171-64
38 10 1.5 130 12 US-09-804-980-64
39 10 1.5 130 12 US-10-282-122A-61922
40 10 1.5 130 12 US-10-282-122A-62378
41 10 1.5 130 12 US-10-282-122A-64413
42 10 1.5 130 12 US-09-943-443-64
43 10 1.5 130 14 US-10-138-473-64
44 10 1.5 130 16 US-10-389-566-1286
45 10 1.5 132 16 US-10-389-566-1794

ALIGNMENTS

RESULT 1

US-09-864-761-36456
; Sequence 36456, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharon G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; FILE REFERENCE: Aeomica-X-1
; CURRENT APPLICATION NUMBER: US/09/864, 761
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30

```

; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Anomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 36456
; LENGTH: 208
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AC004827.1
; OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 5.7
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 4.1
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 5.4
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 3.6
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 7.7
; OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 5.5
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 5.8
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 8.4
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 1.9
; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 4.5
; OTHER INFORMATION: EST_HUMAN HIT: AI736861.1, EVALUATE 2.00e-07
US-09-864-761-36456

Query Match      8.1%; Score 55; DB 9; Length 208;
Best Local Similarity 100.0%; Pred. No. 1e-39;
Matches 55; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 MPRAQSSASVOPVADPFAIVSRQAQWVEILSDENRNLROELGCGYKVARLQK 55
DB      154 MPRAQSSASVOPVADPFAIVSRQAQWVEILSDENRNLROELGCGYKVARLQK 208

RESULT 2
US-09-864-761-39389
; Sequence 39389, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharron G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; TITLE OF INVENTION: GENE EXPRESSION ANALYSIS BY MICROARRAY
; FILE REFERENCE: Aecomica-X-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669

```

```

; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Anomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 39389
; LENGTH: 48
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AC004827.1
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 8.2
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 10
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 7.5
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 9.6
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 9
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 8.2
; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 7.8
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 8
; OTHER INFORMATION: EST_HUMAN HIT: BF032018.1, EVALUATE 8.00e-20
; OTHER INFORMATION: SWISSPROT HIT: Q99972, EVALUATE 9.40e-01
US-09-864-761-39389

Query Match      7.1%; Score 48; DB 9; Length 48;
Best Local Similarity 100.0%; Pred. No. 3.7e-34;
Matches 48; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      56 VETETQRVSEAYENLVKSSKREALEKAMRNKLRGEIRRMHDFNRDLR 103
DB      1 VETETQRVSEAYENLVKSSKREALEKAMRNKLRGEIRRMHDFNRDLR 48

RESULT 3
US-09-864-761-43939
; Sequence 43939, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharron G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; TITLE OF INVENTION: GENE EXPRESSION ANALYSIS BY MICROARRAY
; FILE REFERENCE: Aecomica-X-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666

```

PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00667
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00664
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00669
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00665
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00668
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00663
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00662
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00661
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00670
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: US 60/234,687
PRIOR FILING DATE: 2000-09-21
PRIOR APPLICATION NUMBER: US 09/608,408
PRIOR FILING DATE: 2000-06-30
PRIOR APPLICATION NUMBER: US 09/774,203
PRIOR FILING DATE: 2001-01-29
NUMBER OF SEQ ID NOS: 49117
SOFTWARE: Annotax Sequence Listing Engine vers. 1.1
SEQ ID NO 43939
LENGTH: 26
TYPE: PRT
ORGANISM: Homo sapiens
FEATURE:
OTHER INFORMATION: MAP TO AC004827.1
OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 0.56
OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 0.59
OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 0.64
OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 0.71
OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 0.7
US-09-864-761-43939
Query Match 3.8%; Score 26; DB 9; Length 26;
Best Local Similarity 100.0%; Pred. No. 4.1e-15;
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Y 650 SNTLKRTPQILGQBPDAEMVEYLI 675
b 1 SNTLKRTPQILGQBPDAEMVEYLI 26
RESULT 4
US-10-204-887-87
Sequence 87, Application US/10204887
Publication No. US20030124569A1
GENERAL INFORMATION:
APPLICANT: INCYTE GENOMICS, INC.
APPLICANT: PANZER, Scott R.
APPLICANT: SPIRO, Peter A.
APPLICANT: BANVILLE, Steven C.
APPLICANT: SHAH, Purvi
APPLICANT: CEALUP, Michael S.
APPLICANT: CHENG, Simon C.
APPLICANT: CHEN, Alice
APPLICANT: D'SA, Steven A.
APPLICANT: AMSHEY, Stefan
APPLICANT: DAHL, Christopher R.
APPLICANT: DAM, Tam C.
APPLICANT: DANIELS, Susan E.
APPLICANT: DUFOUR, Gerard E.
APPLICANT: FLORES, Vincent
APPLICANT: FONG, Willy T.
APPLICANT: GREENAWALT, Lila B.
APPLICANT: HILLMAN, Jennifer L.
APPLICANT: JONES, Anissa L.
APPLICANT: LIU, Tommy P.
APPLICANT: ROSEBERRY, Ann M.
APPLICANT: ROSEN, Bruce H.
APPLICANT: RUSSO, Frank D.
APPLICANT: STOCKREHER, Theresa K.
APPLICANT: DAFFO, Abel
APPLICANT: WRIGHT, Rachel J.
APPLICANT: YAP, Pierre E.
APPLICANT: YU, Jimmy Y.
APPLICANT: BRADLEY, Diana L.
APPLICANT: BRATCHER, Shawn R.
APPLICANT: CHEN, Wensheng
APPLICANT: COHEN, Howard J.
APPLICANT: HODGSON, David M.
APPLICANT: LINCOLN, Stephen E.
TITLE OF INVENTION: SECRETORY MOLECULES
FILE REFERENCE: PT-1134.PCT
CURRENT APPLICATION NUMBER: US/10/204,887
CURRENT FILING DATE: 2002-08-21
PRIOR APPLICATION NUMBER: 60/205,324; 60/205,286
PRIOR FILING DATE: 2000-02-24; 2000-05-16; 2000-05-17; 2000-05-17;
2000-05-17; 2000-05-17
NUMBER OF SEQ ID NOS: 159
SOFTWARE: PERL Program
SEQ ID NO 87
LENGTH: 608
TYPE: PRT
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc feature
OTHER INFORMATION: Incyte ID No. US20030124569A1 LI:212029.1.orf2.2000FEB01
FEATURE:
NAME/KEY: unsure
LOCATION: 18, 388
OTHER INFORMATION: unknown or other
US-10-204-887-87
Query Match 3.8%; Score 24; DB 14; Length 608;
Best Local Similarity 100.0%; Pred. No. 4.2e-12;
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 256 REKEERILALEADMTKWEQKYLEE 279
Db 323 REKEERILALEADMTKWEQKYLEE 346
RESULT 5
US-10-298-417-2
Sequence 2, Application US/10298417
Publication No. US20030124603A1
GENERAL INFORMATION:
APPLICANT: Miyuki Nishimura
APPLICANT: Mayumi Asano
APPLICANT: Yuichi Ono
APPLICANT: Koji Morimoto
APPLICANT: Masakazu Takeuchi
APPLICANT: Yoko Inoue
APPLICANT: Toshio Imai
APPLICANT: Yoshimi Takai
TITLE OF INVENTION: Exocrine gland tight junction-constituting protein JPAP family
FILE REFERENCE:
CURRENT APPLICATION NUMBER: US/10/298,417
CURRENT FILING DATE: 2003-01-22
PRIOR APPLICATION NUMBER: JP 2001-352241
PRIOR FILING DATE: 2001-11-16
NUMBER OF SEQ ID NOS: 10
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 2
LENGTH: 882
TYPE: PRT
ORGANISM: Mus musculus
US-10-298-417-2

```
Query Match
Best Local Similarity 100.0%; Score 24; DB 14; Length 882;
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 256 REKEERILALEADMTKWEQKYLEE 279
      |||||
Db 596 REKEERILALEADMTKWEQKYLEE 619
      |||||

RESULT 6
US-10-094-466-52
; Sequence 52, Application US/10094466
; Publication No. US20030203363A1
; GENERAL INFORMATION:
; APPLICANT: Spyttek et al.
; TITLE OF INVENTION: NOVEL HUMAN PROTEINS, POLYNUCLEOTIDES ENCODING THEM
; TITLE OF INVENTION: AND METHODS OF USING
; FILE REFERENCE: THE SAME
; CURRENT APPLICATION NUMBER: US/10/094,466
; CURRENT FILING DATE: 2002-03-07
; PRIOR APPLICATION NUMBER: 60/274,281
; PRIOR FILING DATE: 2001-03-08
; PRIOR APPLICATION NUMBER: 60/288,148
; PRIOR FILING DATE: 2001-05-02
; PRIOR APPLICATION NUMBER: 60/274,849
; PRIOR FILING DATE: 2001-03-09
; PRIOR APPLICATION NUMBER: 60/275,235
; PRIOR FILING DATE: 2001-03-12
; PRIOR APPLICATION NUMBER: 60/338,375
; PRIOR FILING DATE: 2001-12-04
; PRIOR APPLICATION NUMBER: 60/275,579
; PRIOR FILING DATE: 2001-03-13
; PRIOR APPLICATION NUMBER: 60/335,302
; PRIOR FILING DATE: 2001-10-31
; PRIOR APPLICATION NUMBER: 60/275,601
; PRIOR FILING DATE: 2001-03-13
; PRIOR APPLICATION NUMBER: 60/276,000
; PRIOR FILING DATE: 2001-03-14
; PRIOR APPLICATION NUMBER: 60/277,338
; PRIOR FILING DATE: 2001-03-20
; NUMBER OF SEQ ID NOS: 114
; SOFTWARE: PatIn 2.1
; SEQ ID NO 52
; LENGTH: 467
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-094-466-52

Query Match
Best Local Similarity 100.0%; Score 18; DB 15; Length 467;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 262 ILALEADMTKWEQKYLEE 279
      |||||
Db 246 ILALEADMTKWEQKYLEE 263
      |||||

RESULT 7
US-10-298-417-4
; Sequence 4, Application US/10298417
; Publication No. US20030124603A1
; GENERAL INFORMATION:
; APPLICANT: Miyuki Nishimura
; APPLICANT: Mayumi Asano
; APPLICANT: Yuichi Ono
; APPLICANT: Koji Morimoto
; APPLICANT: Masakazu Takeuchi
; APPLICANT: Yoko Inoue
; APPLICANT: Toshio Imai
; APPLICANT: Yoshimi Takai
; TITLE OF INVENTION: Exocrine gland tight junction-constituting protein JPAP family
; FILE REFERENCE:
; CURRENT APPLICATION NUMBER: US/10/298,417
; CURRENT FILING DATE: 2003-01-22
; PRIOR APPLICATION NUMBER: JP 2001-352241
; PRIOR FILING DATE: 2001-11-16
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 4
; LENGTH: 772
; TYPE: PRT
; ORGANISM: Mus musculus
US-10-298-417-4

Query Match
Best Local Similarity 100.0%; Score 18; DB 14; Length 772;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 262 ILALEADMTKWEQKYLEE 279
      |||||
Db 556 ILALEADMTKWEQKYLEE 573
      |||||

RESULT 8
US-10-243-552-605
; Sequence 605, Application US/10243552
; Publication No. US20030224379A1
; GENERAL INFORMATION:
; APPLICANT: Tang, Y. Tom
; APPLICANT: Yang, Yonghong
; APPLICANT: Wang, Zhiwei
; APPLICANT: Wang, Gezhi
; APPLICANT: Ma, Yunqing
; TITLE OF INVENTION: Novel Nucleic Acids and
; TITLE OF INVENTION: Polypeptides
; FILE REFERENCE: 807A
; CURRENT APPLICATION NUMBER: US/10/243,552
; CURRENT FILING DATE: 2002-09-12
; PRIOR APPLICATION NUMBER: US 60/322,511
; PRIOR FILING DATE: 2001-09-13
; PRIOR APPLICATION NUMBER: PCT/US00/35017
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: US 09/488,725
; PRIOR FILING DATE: 2000-01-21
; PRIOR APPLICATION NUMBER: US 09/552,317
; PRIOR FILING DATE: 2000-04-25
; PRIOR APPLICATION NUMBER: PCT/US01/02623
; PRIOR FILING DATE: 2001-01-25
; PRIOR APPLICATION NUMBER: US 09/491,404
; PRIOR FILING DATE: 2000-01-25
; PRIOR APPLICATION NUMBER: PCT/US01/03800
; PRIOR FILING DATE: 2001-02-05
; PRIOR APPLICATION NUMBER: US 09/496,914
; PRIOR FILING DATE: 2000-02-03
; PRIOR APPLICATION NUMBER: US 09/560,875
; PRIOR FILING DATE: 2000-04-27
; PRIOR APPLICATION NUMBER: PCT/US01/04927
; PRIOR FILING DATE: 2001-02-26
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 998
; SOFTWARE: PC_FL_Genes Version 5.0
; SEQ ID NO 605
; LENGTH: 228
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-243-552-605

Query Match
Best Local Similarity 100.0%; Score 11; DB 12; Length 228;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 507 AAAAATAAAAAA 517
      |||||
```

78 AAAAAA 88

RESULT 9

US-10-369-493-13871
Sequence 13871, Application US/10369493
Publication No. US20030233675A1

GENERAL INFORMATION:

APPLICANT: Cao, Yongwei
APPLICANT: Hinkle, Gregory J.
APPLICANT: Slater, Steven C.
APPLICANT: Goldman, Barry S.
APPLICANT: Chen, Xianfeng

TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF

FILE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES

FILE REFERENCE: 38-10(52052)B

CURRENT APPLICATION NUMBER: US/10/369,493

CURRENT FILING DATE: 2003-02-28

PRIOR APPLICATION NUMBER: US 60/360,039

PRIOR FILING DATE: 2002-02-21

NUMBER OF SEQ ID NOS: 47374

SEQ ID NO 13871

LENGTH: 407

TYPE: PRT

ORGANISM: Pseudomonas fluorescens

US-10-369-493-13871

Query Match 1.6%; Score 11; DB 15; Length 407;
Best Local Similarity 100.0%; Pred. No. 0.73;

Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Y 507 AAAAAA 517

|||||

87 AAAAAA 97

RESULT 10

US-10-369-493-10023
Sequence 10023, Application US/10369493
Publication No. US20030233675A1

GENERAL INFORMATION:

APPLICANT: Cao, Yongwei
APPLICANT: Hinkle, Gregory J.
APPLICANT: Slater, Steven C.
APPLICANT: Goldman, Barry S.
APPLICANT: Chen, Xianfeng

TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF

FILE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES

FILE REFERENCE: 38-10(52052)B

CURRENT APPLICATION NUMBER: US/10/369,493

CURRENT FILING DATE: 2003-02-28

PRIOR APPLICATION NUMBER: US 60/360,039

PRIOR FILING DATE: 2002-02-21

NUMBER OF SEQ ID NOS: 47374

SEQ ID NO 10023

LENGTH: 443

TYPE: PRT

ORGANISM: magnetite-containing magnetic coccus

US-10-369-493-10023

Query Match 1.6%; Score 11; DB 15; Length 443;
Best Local Similarity 100.0%; Pred. No. 0.78;

Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Y 602 APTAPTPTCA 612

|||||

180 APTAPTPTCA 190

RESULT 11

US-10-369-493-3424
Sequence 3424, Application US/10369493
Publication No. US20030233675A1

GENERAL INFORMATION:

APPLICANT: Cao, Yongwei
APPLICANT: Hinkle, Gregory J.
APPLICANT: Slater, Steven C.
APPLICANT: Goldman, Barry S.
APPLICANT: Chen, Xianfeng

TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF

FILE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES

FILE REFERENCE: 38-10(52052)B

CURRENT APPLICATION NUMBER: US/10/369,493

CURRENT FILING DATE: 2003-02-28

PRIOR APPLICATION NUMBER: US 60/360,039

PRIOR FILING DATE: 2002-02-21

NUMBER OF SEQ ID NOS: 47374

SEQ ID NO 3424

LENGTH: 1906

TYPE: PRT

ORGANISM: Neurospora crassa

FEATURE:

NAME/KEY: unsure

LOCATION: (1)..(1906)

OTHER INFORMATION: unsure at all Xaa locations

US-10-369-493-3424

Query Match 1.6%; Score 11; DB 15; Length 1906;
Best Local Similarity 100.0%; Pred. No. 3;

Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 507 AAAAAA 517

|||||

114 AAAAAA 124

RESULT 12

US-10-282-122A-60991

Sequence 60991, Application US/10282122A

Publication No. US20040029129A1

GENERAL INFORMATION:

APPLICANT: Wang, Liangsu
APPLICANT: Zamudio, Carlos
APPLICANT: Malone, Cheryl
APPLICANT: Haselbeck, Robert
APPLICANT: Ohlsen, Kari
APPLICANT: Zyskind, Judith
APPLICANT: Wall, Daniel
APPLICANT: Trawick, John

APPLICANT: Carr, Grant

APPLICANT: Yamamoto, Robert

APPLICANT: Forsyth, R.

APPLICANT: Xu, H.

TITLE OF INVENTION: Identification of Essential Genes in Microorganisms

FILE REFERENCE: ELITRA.034A

CURRENT APPLICATION NUMBER: US/10/282,122A

CURRENT FILING DATE: 2003-02-20

PRIOR APPLICATION NUMBER: 60/191,078

PRIOR FILING DATE: 2000-03-21

PRIOR APPLICATION NUMBER: 60/206,848

PRIOR FILING DATE: 2000-05-23

PRIOR APPLICATION NUMBER: 60/207,727

PRIOR FILING DATE: 2000-05-26

PRIOR APPLICATION NUMBER: 60/230,335

PRIOR FILING DATE: 2000-09-06

PRIOR APPLICATION NUMBER: 60/230,347

PRIOR FILING DATE: 2000-09-09

PRIOR APPLICATION NUMBER: 60/242,578

PRIOR FILING DATE: 2000-10-23

PRIOR APPLICATION NUMBER: 60/253,625

PRIOR FILING DATE: 2000-11-27

PRIOR APPLICATION NUMBER: 60/257,931

PRIOR FILING DATE: 2000-12-22

PRIOR APPLICATION NUMBER: 60/267,636

PRIOR FILING DATE: 2001-02-09

PRIOR APPLICATION NUMBER: 60/269,308

; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 60991
; LENGTH: 120
; TYPE: PRT
; ORGANISM: Listeria monocytogenes
US-10-282-122A-60991

Query Match 1.5%; Score 10; DB 12; Length 120;
Best Local Similarity 100.0%; Pred. No. 1.8;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 500 AAAPVAVAAA 509
| | | | |
Db 35 AAAPVAVAAA 44

RESULT 13
US-10-389-566-1360
; Sequence 1360, Application US/10389566
; Publication No. US20040025202A1
; GENERAL INFORMATION:
; APPLICANT: Monsanto Technology, LLC
; APPLICANT: Laurie, Cathy C
; TITLE OF INVENTION: Nucleic Acid Molecules Associated with Oil in Plants
; FILE REFERENCE: 38-77(52900)D
; CURRENT APPLICATION NUMBER: US/10/389,566
; CURRENT FILING DATE: 2003-03-31
; PRIOR APPLICATION NUMBER: US 60/365,301
; PRIOR FILING DATE: 2002-03-15
; PRIOR APPLICATION NUMBER: US 60/391,786
; PRIOR FILING DATE: 2002-06-25
; PRIOR APPLICATION NUMBER: US 60/392,018
; PRIOR FILING DATE: 2002-06-26
; NUMBER OF SEQ ID NOS: 2459
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 1360
; LENGTH: 120
; TYPE: PRT
; ORGANISM: Listeria innocua
US-10-389-566-1360

Query Match 1.5%; Score 10; DB 16; Length 120;
Best Local Similarity 100.0%; Pred. No. 1.8;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 500 AAAPVAVAAA 509
| | | | |
Db 35 AAAPVAVAAA 44

RESULT 14
US-10-389-566-1545
; Sequence 1545, Application US/10389566
; Publication No. US20040025202A1
; GENERAL INFORMATION:
; APPLICANT: Monsanto Technology, LLC
; APPLICANT: Laurie, Cathy C
; TITLE OF INVENTION: Nucleic Acid Molecules Associated with Oil in Plants
; FILE REFERENCE: 38-77(52900)D
; CURRENT APPLICATION NUMBER: US/10/389,566
; CURRENT FILING DATE: 2003-03-31
; PRIOR APPLICATION NUMBER: US 60/365,301
; PRIOR FILING DATE: 2002-03-15
; PRIOR APPLICATION NUMBER: US 60/391,786
; PRIOR FILING DATE: 2002-06-25
; PRIOR APPLICATION NUMBER: US 60/392,018
; PRIOR FILING DATE: 2002-06-26
; NUMBER OF SEQ ID NOS: 2459
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 1545

; LENGTH: 120
; TYPE: PRT
; ORGANISM: Listeria monocytogenes
US-10-389-566-1545

Query Match 1.5%; Score 10; DB 16; Length 120;
Best Local Similarity 100.0%; Pred. No. 1.8;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 500 AAAPVAVAAA 509
| | | | |
Db 35 AAAPVAVAAA 44

RESULT 15
US-09-815-242-13247
; Sequence 13247, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Cair, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 13247
; LENGTH: 122
; TYPE: PRT
; ORGANISM: Streptococcus pneumoniae
US-09-815-242-13247

Query Match 1.5%; Score 10; DB 9; Length 122;
Best Local Similarity 100.0%; Pred. No. 1.8;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 500 AAAPVAVAAA 509
| | | | |
Db 35 AAAPVAVAAA 44

Search completed: June 16, 2004, 19:33:18
Job time : 50 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

* protein - protein search, using sw model

run on: June 16, 2004, 19:23:53 ; Search time 20 Seconds
(without alignments)
3246.464 Million cell updates/sec

File: US-09-332-063-2
Perfect score: 675
Sequence: 1 MPRAQSSASVQVPADPEA.....KTPILQILQEPDAEMVEYLI 675

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 283366 seqs, 96191526 residues

Word size : 0

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : PIR 78.*

1: Piri.*

2: Piri2.*

3: Piri3.*

4: Piri4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

result No.	Score	Query Match	Length	ID	Description
1	12	1.8	172	2 S35568	sex-determining pr
2	11	1.6	109	1 R6UTP1	acidic ribosomal p
3	11	1.6	270	2 E87649	hypothetical prote
4	11	1.6	312	2 T21351	hypothetical prote
5	11	1.6	553	2 B55514	dihydrolipoamide S
6	10	1.5	82	1 FDPLAW	antifreeze protein
7	10	1.5	82	2 S02326	antifreeze protein
8	10	1.5	82	2 A05161	antifreeze protein
9	10	1.5	82	2 I51125	antifreeze protein
10	10	1.5	101	2 H69091	ribosomal protein
11	10	1.5	106	1 R6DQ2	acidic ribosomal p
12	10	1.5	120	2 AD1106	ribosomal protein
13	10	1.5	120	2 AD1468	ribosomal protein
14	10	1.5	122	2 C95157	ribosomal protein
15	10	1.5	124	2 I40348	ribosomal protein
16	10	1.5	124	2 I40350	ribosomal protein
17	10	1.5	124	2 AF3345	LSU ribosomal prot
18	10	1.5	125	2 AC2817	50S ribosomal prot
19	10	1.5	125	2 D97595	ribosomal protein
20	10	1.5	127	2 H87310	ribosomal protein
21	10	1.5	127	2 B98023	50S ribosomal prot
22	10	1.5	128	1 R7H612	ribosomal protein
23	10	1.5	130	2 A70615	probable ribosomal
24	10	1.5	130	2 S41123	ribosomal protein
25	10	1.5	136	2 A56062	Alu RNA-binding pr
26	10	1.5	136	2 S34196	signal recognition
27	10	1.5	156	1 A49342	acetyl-CoA carboxy
28	10	1.5	158	2 P87420	hypothetical prote
29	10	1.5	308	2 A38562	pollen allergen pI

30	10	1.5	315	2 B99607	hypothetical prote
31	10	1.5	635	2 C82079	pyruvate dehydroge
32	10	1.5	747	1 EABO	elastin precursor,
33	10	1.5	762	2 H87302	chemotaxis protein
34	10	1.5	850	2 JC5047	ras GTPase-activat
35	10	1.5	1887	2 S61703	fatty-acid synthas
36	9	1.3	102	2 D64363	ribosomal protein
37	9	1.3	106	2 C69436	LSU ribosomal prot
38	9	1.3	110	2 T37490	ribosomal protein
39	9	1.3	111	2 E72524	probable ribosomal
40	9	1.3	135	2 S34952	histone H1 - Rupo
41	9	1.3	140	2 D97722	50S ribosomal prot
42	9	1.3	149	2 AB0770	probable protein-t
43	9	1.3	191	2 H83151	hypothetical prote
44	9	1.3	200	2 G84822	probable LIM-domai
45	9	1.3	227	2 G70555	hypothetical prote

ALIGNMENTS

RESULT 1

S35568
sex-determining protein Sry - multimammate rat (Mastomys hildebrandtii)
C:Species: Mastomys hildebrandtii
C:Date: 19-Mar-1997 #sequence_revision 19-Mar-1997 #text_change 24-Sep-1999
C:Accession: S35568
R:Tucker, P.K.; Lundrigan, B.L.
Nature 364, 715-717, 1993
A:Title: Rapid evolution of the sex determining locus in Old World mice and rats.
A:Reference number: S35565; MUID:93361118; PMID:8355784
A:Accession: S35568
A:Molecule type: DNA
A:Residues: 1-172 <TUC>
A:Cross-references: GB:L29542; NID:G496161; PIDN:AAA40587.1; PID:G496162
C:Genetics:
A:Gene: Sry
A:Map position: Y
C:Superfamily: unassigned HMG box proteins; HMG box homology
C:Keywords: DNA binding
P:2-77/Domain: HMG box homology <HMG1>

Query Match		1.8%	Score 12;	DB 2;	Length 172;
Best Local Similarity		100.0%;	Pred. No. 0.017;		
Matches 12;		Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
Qy	507	AAAAAPAAAAAP	518		
Db	153	AAAAAPAAAAAP	164		

RESULT 2

R6UTP1
acidic ribosomal protein P1 - Trypanosoma cruzi
C:Species: Trypanosoma cruzi
C:Date: 31-Dec-1992 #sequence_revision 31-Dec-1992 #text_change 22-Jun-1999
C:Accession: S22644; S15924
R:Vazquez, M.P.; Schijman, A.G.; Levin, M.J.
Nucleic Acids Res. 20, 2599, 1992
A:Title: Nucleotide sequence of a cDNA encoding a Trypanosoma cruzi acidic ribosomal P1
A:Reference number: S22644; MUID:92285148; PMID:1598221
A:Accession: S22644
A:Status: nucleic acid sequence not shown
A:Molecule type: mRNA
A:Residues: 1-109 <VAZ>
A:Cross-references: EMBL:X55025; NID:G10629; PIDN:CAA46159.1; PID:G10630
C:Superfamily: rat acidic ribosomal protein P1
C:Keywords: phosphoprotein; protein biosynthesis; ribosome
Query Match
Best Local Similarity 100.0%;

Query Match		1.6%	Score 11;	DB 1;	Length 109;
Best Local Similarity		100.0%;	Pred. No. 0.081;		
Matches 11;		Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;

2Y 507 AAAAAPAAAAA 517
| | | | | | | | | |
Db 82 AAAAAPAAAAA 92

RESULT 3
387649
;Species: Caulobacter crescentus
;Date: 20-Apr-2001 #sequence_revision 20-Apr-2001 #text_change 20-Apr-2001
;Accession: B87649
;Nierman, W.C.; Feldblyum, T.V.; Paulsen, I.T.; Nelson, K.E.; Eisen, J.; Heidelberg, J.
B.; Laub, M.T.; DeBoy, R.T.; Dodson, R.J.; Durkin, A.S.; Gwinn, M.L.; Haft, D.H.; Kolon
J.; Ermolaeva, M.; White, O.; Salzberg, S.L.; Shapiro, L.; Venter, J.C.; Fraser, C.M.
Proc. Natl. Acad. Sci. U.S.A. 98, 4136-4141, 2001
;Title: Complete Genome Sequence of Caulobacter crescentus.
;Reference number: A87249; MUID:21173698; PMID:11259847
;Accession: B87649
;Status: preliminary
;Molecule type: DNA
;Residues: 1-270 <STO>
;Cross-references: GB:AE005673; MID:g13424915; PIDN:AAK25193.1; GSPDB:GN00148
;Genetics:
;Gene: CC3231

Query Match 1.6%; Score 11; DB 2; Length 270;
Best Local Similarity 100.0%; Pred. No. 0.17;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

2Y 602 APTAPPTPTPA 612
| | | | | | | | | |
Db 84 APTAPPTPTPA 94

RESULT 4
721351
;Species: Caenorhabditis elegans
;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 11-Jan-2000
;Accession: T21351
;Submitted to the EMBL Data Library, September 1996
;Reference number: Z19411
;Accession: T21351
;Status: preliminary; translated from GB/EMBL/DDBU
;Molecule type: DNA
;Residues: 1-312 <WIL>
;Cross-references: EMBL:Z79754; PIDN:CAB02098.1; GSPDB:GN00019; CESP:F25H2.10
;Experimental source: clone F25H2
;Genetics:
;Gene: CESP:F25H2.10
;Map position: 1
;Introns: 49/1; 157/3
;Superfamily: rat acidic ribosomal protein P0

Query Match 1.6%; Score 11; DB 2; Length 312;
Best Local Similarity 100.0%; Pred. No. 0.2;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

2Y 508 AAAAAPAAAAA 518
| | | | | | | | | |
Db 276 AAAAAPAAAAA 286

RESULT 5
35514
;Species: Alcaligenes eutrophus
;Date: 18-Aug-1995 #sequence_revision 18-Aug-1995 #text_change 05-May-2000
;Accession: B5514
;Hein, S.; Steinbuechel, A.
;Bacteriol. 176, 4394-4408, 1994
;Title: Biochemical and molecular characterization of the Alcaligenes eutrophus pyruvat

A:Reference number: A55514; MUID:94292470; PMID:8021225
A:Accession: B5514
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-553 <HEI>
A:Cross-references: GB:U09865; MID:g497263; PIDN:AAA21599.1; PID:g497265
A:Genetics:
A:Gene: pdhE
C:Superfamily: dihydrolipoamide acetyltransferase; lipoyl/biotin-binding homology
F:6-78/Domain: lipoyl/biotin-binding homology <LPI>
F:124-196/Domain: lipoyl/biotin-binding homology <LPI>
F:526,530/Active site: His, Asp #status predicted
Query Match 1.6%; Score 11; DB 2; Length 553;
Best Local Similarity 100.0%; Pred. No. 0.32;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 507 AAAAAPAAAAA 517
| | | | | | | | | |
Db 199 AAAAAPAAAAA 209

RESULT 6
FDLAW
antifreeze protein A precursor - winter flounder
C:Species: Pseudopleuronectes americanus (winter flounder)
C:Date: 13-Aug-1986 #sequence_revision 13-Aug-1986 #text_change 22-Jun-1999
C:Accession: J50704; A03194
R: Davies, P.L.
Gene 112, 163-170, 1992
A:Title: Conservation of antifreeze protein-encoding genes in tandem repeats.
A:Reference number: JH0627; MUID:92209995; PMID:1555765
A:Accession: J50704
A:Molecule type: DNA
A:Residues: 1-82 <DAL>
A:Cross-references: GB:M62412; GB:M62416; MID:g213592; PIDN:AAA49471.1; PID:g213593
R: Davies, P.L.; Roach, A.H.; Hew, C.L.
Proc. Natl. Acad. Sci. U.S.A. 79, 335-339, 1982
A:Title: DNA sequence coding for an antifreeze protein precursor from winter flounder.
A:Reference number: A03194; MUID:82197490; PMID:6952188
A:Accession: A03194
A:Molecule type: mRNA
A:Residues: 1-82 <DA2>
A:Experimental source: clones 4-2b and 2A-7c
A:Note: the authors translated the codon AGC for residue 24 as Arg
C:Genetics:
A:Introns: 19/2
C:Superfamily: antifreeze protein
C:Keywords: antifreeze; plasma; tandem repeat
F:1-21/Domain: signal sequence #status predicted <SIG>
F:22-44/Domain: propeptide #status predicted <PRO>
F:45-82/Product: antifreeze protein A #status predicted <MAT>

Query Match 1.5%; Score 10; DB 1; Length 82;
Best Local Similarity 100.0%; Pred. No. 0.46;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 507 AAAAAPAAAAA 516
| | | | | | | | | |
Db 34 AAAAAPAAAAA 43

RESULT 7
S02326
antifreeze protein A - winter flounder
C:Species: Pseudopleuronectes americanus (winter flounder)
C:Date: 30-Sep-1989 #sequence_revision 30-Sep-1989 #text_change 24-Oct-2000
C:Accession: S02326; JH0627
R: Scott, G.K.; Davies, P.L.; Kao, M.H.; Fletcher, G.L.
J. Mol. Evol. 27, 29-35, 1988
A:Title: Differential amplification of antifreeze protein genes in the Pleuronectinae.
A:Reference number: S02326; MUID:88259236; PMID:313486

Accession: S02326
Molecule type: DNA
Residues: 1-82 <SCO>
Cross-references: EMBL:X07506; NID:964211; PIDN:CAA30389.1; PID:964212
Davies, P.L.
ene 112, 163-170, 1992
Title: Conservation of antifreeze protein-encoding genes in tandem repeats.
Reference number: JH0627; MUID:92209995; PMID:1555765
Accession: JH0627
Molecule type: DNA
Residues: 1-82 <DAV>
Cross-references: GB:M62415
Genetics:
Introns: 19/2
Superfamily: antifreeze protein
Keywords: antifreeze

Query Match 1.5%; Score 10; DB 2; Length 82;
Best Local Similarity 100.0%; Pred. No. 0.46;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Y 507 AAAAAAPAAAA 516
|||||
b 34 AAAAAAPAAAA 43

RESULT 8
05161
ntifreeze protein B precursor - winter flounder
Species: Pseudopleuronectes americanus (winter flounder)
Date: 05-Jun-1987 #sequence_revision 05-Jun-1987 #text_change 24-Oct-2000
Accession: A05161
Davies, P.L.; Hough, C.; Scott, G.K.; Ng, N.; White, B.N.; Hew, C.L.
Biochem. J. 259, 9241-9247, 1984
Reference number: A05161; MUID:84264559; PMID:6086629
Accession: A05161
Molecule type: DNA
Residues: 1-82 <DAV>
Cross-references: GB:L00138; GB:J00929; NID:G343126; PIDN:AA859964.1; PID:9457351
Genetics:
Introns: 19/2
Superfamily: antifreeze protein
Keywords: antifreeze

Query Match 1.5%; Score 10; DB 2; Length 82;
Best Local Similarity 100.0%; Pred. No. 0.46;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Y 507 AAAAAAPAAAA 516
|||||
b 34 AAAAAAPAAAA 43

RESULT 9
51125
ntifreeze protein - winter flounder
Species: Pseudopleuronectes americanus (winter flounder)
Date: 13-Sep-1996 #sequence_revision 13-Sep-1996 #text_change 24-Oct-2000
Accession: I51125
Pickett, M.; Scott, G.; Davies, P.; Wang, N.; Joshi, S.; Hew, C.
J. Biochem. 143, 35-38, 1984
Title: Sequence of an antifreeze protein precursor.
Reference number: I51125; MUID:84285392; PMID:6547905
Accession: I51125
Status: preliminary; translated from GB/EMBL/DBJ
Molecule type: mRNA
Residues: 1-82 <PIC>
Cross-references: GB:M28337; NID:G213581; PIDN:AAA49466.1; PID:G213582
Genetics:
Gene: AFP
Superfamily: antifreeze protein
Keywords: antifreeze

Query Match 1.5%; Score 10; DB 2; Length 82;
Best Local Similarity 100.0%; Pred. No. 0.46;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 507 AAAAAAPAAAA 516
|||||
DB 34 AAAAAAPAAAA 43

RESULT 10
H69091
ribosomal protein Lp1 - Methanobacterium thermoautotrophicum (strain Delta H)
Species: Methanobacterium thermoautotrophicum
C>Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 13-Aug-1999
C:Accession: H69091
R:Smith, D.R.; Doucette-Stamm, L.A.; Deloughery, C.; Lee, H.; Dubois, J.; Adredge, T.;
Qiu, D.; Spadafora, R.; Vicaire, R.; Wang, Y.; Wierzbowski, J.; Gibson, R.; Jiwani, N.
Kl, S.; Church, G.M.; Daniels, C.J.; Mao, J.; Rice, P.; Noelling, J.; Reeve, J.N.
J. Bacteriol. 179, 7135-7155, 1997
A>Title: Complete genome sequence of Methanobacterium thermoautotrophicum Delta H: funct
J>Title: Complete genome sequence of Methanobacterium thermoautotrophicum Delta H: funct
A:Reference number: A69000; MUID:98037514; PMID:9371463
A:Accession: H69091
A>Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-101 <MTH>
A:Cross-references: GB:AE000926; GB:AE000666; NID:G2622806; PIDN:AA86154.1; PID:G262281
A:Experimental source: strain Delta H
C:Genetics:
A:Gene: MTH1682
C:Superfamily: rat acidic ribosomal protein P1

Query Match 1.5%; Score 10; DB 2; Length 101;
Best Local Similarity 100.0%; Pred. No. 0.55;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 509 AAAAAAPAAAA 518
|||||
DB 60 AAAAAAPAAAA 69

RESULT 11
R6DOP2
acidic ribosomal protein P2 - slime mold (Dictyostelium discoideum)
N:Alternate names: ribosomal phosphoprotein P2
C:Species: Dictyostelium discoideum
C>Date: 31-Dec-1992 #sequence_revision 31-Dec-1992 #text_change 22-Jun-1999
C:Accession: S14014
R:Prieto, J.; Candel, E.; Coloma, A.
Nucleic Acids Res. 19, 1341, 1991
A>Title: Nucleotide sequence of a cDNA encoding acidic ribosomal phosphoprotein P2 in Di
A:Reference number: S14014; MUID:91232921; PMID:1840653
A:Accession: S14014
A>Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: mRNA
A:Residues: 1-106 <PRI>
A:Cross-references: EMBL:X56192; NID:G7340; PIDN:CAA39655.1; PID:97341
A>Note: the nucleotide sequence was submitted to the EMBL Data Library, November 1990
A>Note: part of this sequence, including the amino end of the mature protein, was confir
C:Superfamily: rat acidic ribosomal protein P1
C:Keywords: phosphoprotein; protein biosynthesis; ribosome
F:1-106/Product: acidic ribosomal protein P2 #status experimental <MAT>
F:64-85/Region: alanine-rich

Query Match 1.5%; Score 10; DB 1; Length 106;
Best Local Similarity 100.0%; Pred. No. 0.57;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 506 AAAAAAPAAAA 515
|||||
DB 63 AAAAAAPAAAA 72

RESULT 12

```

AD1106
ribosomal protein L12 [imported] - Listeria monocytogenes (strain EGD-e)
C:Species: Listeria monocytogenes
C>Date: 27-Nov-2001 #sequence_revision 27-Nov-2001 #text_change 14-Dec-2001
C:Accession: AD1106
R:Glaser, P.; Frangeul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bloeker,
; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussurget, O.; Entian, K.D.; Fsihi, H.
D.; Jones, L.M.; Karst, U.
Science 294, 849-852, 2001
A:Authors: Krefte, J.; Kuhn, M.; Kunst, F.; Kurapkat, G.; Madueno, E.; Maitournam, A.; Ma
ok, C.; Schlueter, T.; Simoes, N.; Tierrez, A.; Vazquez-Boland, J.A.; Voss, H.; Wehland,
A:Title: Comparative Genomics of Listeria species.
A:Reference number: AB1077; MUID:21537279; PMID:11679669
A:Accession: AD1106
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-120 <GLA>
A:Cross-references: GB:NC_003210; PIDN:CAD00778.1; PID:gl6409616; GSPDB:GN00177
A:Experimental source: strain EGD-e
C:Genetics:
A:Gene: rplL
C:Superfamily: Escherichia coli ribosomal protein L12

Query Match 1.5%; Score 10; DB 2; Length 120;
Best Local Similarity 100.0%; Pred. No. 0.64;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

ZY 500 AAAPVAVAAA 509
DB 35 AAAPVAVAAA 44

RESULT 13
AD1468
ribosomal protein L12 [imported] - Listeria innocua (strain Clip11262)
C:Species: Listeria innocua
C>Date: 27-Nov-2001 #sequence_revision 27-Nov-2001 #text_change 14-Dec-2001
C:Accession: AD1468
R:Glaser, P.; Frangeul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bloeker,
; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussurget, O.; Entian, K.D.; Fsihi, H.
D.; Jones, L.M.; Karst, U.
Science 294, 849-852, 2001
A:Authors: Krefte, J.; Kuhn, M.; Kunst, F.; Kurapkat, G.; Madueno, E.; Maitournam, A.; Ma
ok, C.; Schlueter, T.; Simoes, N.; Tierrez, A.; Vazquez-Boland, J.A.; Voss, H.; Wehland,
A:Title: Comparative Genomics of Listeria species.
A:Reference number: AB1077; MUID:21537279; PMID:11679669
A:Accession: AD1468
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-120 <GLA>
A:Cross-references: GB:AL592022; PIDN:CAC95516.1; PID:gl6412712; GSPDB:GN00178
A:Experimental source: strain Clip11262
C:Genetics:
A:Gene: rplL
C:Superfamily: Escherichia coli ribosomal protein L12

Query Match 1.5%; Score 10; DB 2; Length 120;
Best Local Similarity 100.0%; Pred. No. 0.64;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

ZY 500 AAAPVAVAAA 509
DB 35 AAAPVAVAAA 44

RESULT 14
95157
ribosomal protein L7/L12 [imported] - Streptococcus pneumoniae (strain TIGR4)
C:Species: Streptococcus pneumoniae
C>Date: 03-Aug-2001 #sequence_revision 03-Aug-2001 #text_change 24-Aug-2001
C:Accession: C95157
R:Tettelin, H.; Nelson, K.E.; Paulsen, I.T.; Eisen, J.A.; Read, T.D.; Peterson, S.; Heid
m, J.D.; Umayam, L.A.; White, O.; Salzberg, S.L.; Lewis, M.R.; Radune, D.; Holtzapfle,
nson, T.; Hickey, E.K.; Holt, I.E.
Science 293, 498-506, 2001
A:Authors: Loftus, B.J.; Yang, F.; Smith, H.O.; Venter, J.C.; Dougherty, B.A.; Morrison
A:Title: Complete Genome Sequence of a virulent isolate of Streptococcus pneumoniae.
A:Reference number: A95000; MUID:21357209; PMID:11463916
A:Accession: C95157
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-122 <KUP>
A:Cross-references: GB:AB005672; PIDN:AAK75452.1; PID:gl4972838; GSPDB:GN00164; TIGR:SP
A:Experimental source: strain TIGR4
C:Genetics:
A:Gene: SPI354
C:Superfamily: Escherichia coli ribosomal protein L12

Query Match 1.5%; Score 10; DB 2; Length 122;
Best Local Similarity 100.0%; Pred. No. 0.64;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

ZY 500 AAAPVAVAAA 509
DB 35 AAAPVAVAAA 44

RESULT 15
I40348
ribosomal protein L7/L12 - Brucella abortus
C:Species: Brucella abortus
C>Date: 12-Aug-1996 #sequence_revision 12-Aug-1996 #text_change 13-Aug-1999
C:Accession: I40348
R:Oliveira, S.C.; Zhu, Y.; Splitter, G.A.
Gene 140, 137-138, 1994
A:Title: Sequences of the rplJL operon containing the L10 and L7/L12 genes from Brucella;
A:Reference number: I40348; MUID:94171071; PMID:8125331
A:Accession: I40348
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-124 <RES>
A:Cross-references: GB:L19101; NID:G304270; PIDN:AAA19863.1; PID:G304271
C:Superfamily: Escherichia coli ribosomal protein L12

Query Match 1.5%; Score 10; DB 2; Length 124;
Best Local Similarity 100.0%; Pred. No. 0.65;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

ZY 500 AAAPVAVAAA 509
DB 34 AAAPVAVAAA 43

Search completed: June 16, 2004, 19:27:50
Job time : 22 secs

```

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

3M protein - protein search, using sw model

Run on: June 16, 2004, 19:17:13 ; Search time 18 Seconds
(without alignments)
1952.631 Million cell updates/sec

Title: US-09-332-063-2
Perfect score: 675
Sequence: 1 MPRAQSSASVQVPADPPA.....KTPQILQGPDAENVYLI 675

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 141681 seqs, 52070155 residues

Word size : 0

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : SwissProt_42.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

result No.	Score	Query Match	Length	DB ID	Description
1	11	1.6	109	RLA1_TRYCR	P26643 trypanosoma
2	11	1.6	114	RLA2_EIMTE	Q96799 eimeria ten
3	11	1.6	311	RLA0_CAHEL	Q93572 caenorhabdi
4	11	1.6	553	ODP2_AUCEL	Q59098 alcaligenes
5	10	1.5	82	ANPA_PSEAM	P04002 pseudopieur
6	10	1.5	101	RL12_METH	P05394 methanococ
7	10	1.5	101	RL12_METH	O52706 methanococ
8	10	1.5	105	RL2_DICDI	P22683 dictyosteli
9	10	1.5	120	RL7_LISIN	Q92424 listeria in
10	10	1.5	120	RL7_LISIN	Q8Yaa3 listeria mo
11	10	1.5	121	RL7_STRPN	P80714 streptococ
12	10	1.5	122	RL7_FUSNN	Q8rh15 fusobacteri
13	10	1.5	124	RL7_BRUME	P41106 bruceella me
14	10	1.5	125	RL7_AGR75	Q8ue07 agrobacteri
15	10	1.5	125	RL7_RHILO	Q8m67 rhizobium 1
16	10	1.5	125	RL7_THETH	Q8vve2 thermus the
17	10	1.5	125	RL7_THETH	Q8r7u5 thermococ
18	10	1.5	126	RL7_RHIME	Q92488 rhizobium m
19	10	1.5	127	RL7_CAUCR	Q9aa07 carlobacter
20	10	1.5	128	RL7_CORGL	Q8nt28 corynebacte
21	10	1.5	128	RL7_THETH	P29396 thermotoga
22	10	1.5	130	RL7_MYCTU	P37381 mycobacteri
23	10	1.5	136	SR14_HUMAN	P37108 homo sapien
24	10	1.5	156	BCCP_PSEAE	P37799 pseudomonas
25	10	1.5	157	CU63_LOCOM	P45584 locusta mig
26	10	1.5	308	MP5A_LOLPR	Q40240 lolium pere
27	10	1.5	314	RLA0_EIMTE	Q96797 eimeria ten
28	10	1.5	322	RLA0_EUPLU	P50345 lupinus lut
29	10	1.5	747	ELS_BOVIN	P04985 bos taurus
30	10	1.5	848	RSQ2_MOUSE	P58069 mus musculu
31	10	1.5	849	1 FAS2_HUMAN	Q15283 homo sapien
32	10	1.5	1887	1 FAS2_HUMAN	P19097 s fatty aci
33	9	1.3	102	RL12_METUA	P54048 methanococc

ALIGNMENTS

RESULT 1

ID	RLA1_TRYCR	STANDARD;	PRT;	109 AA.
AC	P26643;			
DT	01-AUG-1992 (Rel. 23, Created)			
DT	01-AUG-1992 (Rel. 23, Last sequence update)			
DT	01-OCT-1996 (Rel. 34, Last annotation update)			
DE	60S acidic ribosomal protein P1.			
OS	Trypanosoma cruzi.			
OC	Eukaryota; Euzoosoa; Kinetoplastida; Trypanosomatidae; Trypanosoma.			
OX	NCBI_TaxID=5693;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=92285148; PubMed=1598221;			
RA	Vazquez M., Schijman A., Levin M.;			
RT	"Nucleotide sequence of a cDNA encoding a Trypanosoma cruzi acidic			
RT	ribosomal P1 type protein."			
RL	Nucleic Acids Res. 20:2599-2599(1992).			
CC	!- FUNCTION: Plays an important role in the elongation step of			
CC	protein synthesis.			
CC	!- SUBUNIT: P1 and P2 exist as dimers at the large ribosomal subunit.			
CC	!- SIMILARITY: Belongs to the L12P family of ribosomal proteins.			
CC				
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/			
CC	or send an email to license@isb-sib.ch).			
CC				
EMBL	X65025; CAA46159.1;			
PIR	S22644; R6UTP.			
DR	InterPro; IPR001813; Ribosomal_60S.			
DR	InterPro; IPR001859; Ribosomal_P2.			
DR	Pfam; PF00428; 60S ribosomal; 1.			
DR	PRINTS; PR00456; RIBOSOMALP2.			
XW	Ribosomal protein; Phosphorylation.			
SQ	SEQUENCE 109 AA; 10754 MW; D674031C102F70AF CRC64;			

Query Match 1.6%; Score 11; DB 1; Length 109;
Best Local Similarity 100.0%; Pred.No. 0.038;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	507	AAAAAPAAAAA	517
DB	82	AAAAAPAAAAA	92

RESULT 2

ID	RLA2_EIMTE	STANDARD;	PRT;	114 AA.
AC	Q96799;			
DT	28-FEB-2003 (Rel. 41, Created)			
DT	28-FEB-2003 (Rel. 41, Last sequence update)			
DT	28-FEB-2003 (Rel. 41, Last annotation update)			

DE 60S acidic ribosomal protein P2.
OS Eimeria tenella.
OC Eukaryota; Alveolata; Apicomplexa; Coccidia; Eimeriidae; Eimeriidae;
OC Eimeria.
OX NCBI_TaxID=5802;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=PAPE38;
RA Labbe M., Pery P.;
RT "Molecular cloning of a cDNA encoding an acidic ribosomal protein P2
of Eimeria tenella."
RL Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: Plays an important role in the elongation step of
protein synthesis (By similarity).
CC -!- SUBUNIT: P1 and P2 exist as dimers at the large ribosomal subunit
(By similarity).
CC -!- PTM: Phosphorylated (By similarity).
CC -!- SIMILARITY: Belongs to the L12P family of ribosomal proteins.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).
CC
EMBL; AF353514; AAK38885.1; ALT INIT.
DR InterPro: IPR001813; Ribosomal_60S.
DR Pfam: PF00428; 60S-ribosomal; 1
KW Ribosomal protein; Phosphorylation.
SQ SEQUENCE 114 AA; 11444 MW; 4C083C569078AA9 CRC64;

Query Match 1.6%; Score 11; DB 1; Length 114;
Best Local Similarity 100.0%; Pred. No. 0.039;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 507 AAAAPAAAAA 517
Db 71 AAAAPAAAAA 81

RESULT 3
RLAO CAEEL STANDARD; PRT; 311 AA.
AC Q93572;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE 60S acidic ribosomal protein P0.
GN RPA-C OR P25H2.10.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RA Wilkinson J.;
RL Submitted (NOV-1996) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE OF 1-15.
RC STRAIN=Bristol N2;
RX MEDLINE=97295299; PubMed=9150941;
RA Bini L., Heid H., Liberatori S., Geier G., Pallini V., Zwilling R.;
RT "Two-dimensional gel electrophoresis of Caenorhabditis elegans
homogenates and identification of protein spots by microsequencing."
RL Electrophoresis 18:557-562(1997).
CC -!- FUNCTION: Ribosomal protein P0 is the functional equivalent of
E.coli protein L10.
CC -!- SUBUNIT: P0 forms a pentameric complex by interaction with dimers
of P1 and P2.
CC -!- SIMILARITY: Belongs to the L10P family of ribosomal proteins.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).
CC
EMBL; Z79754; CAS02098.1; -
DR PIR; T21351; T21351.
DR Sienna-2DPAGE; Q93572; -
DR WormPep; P25H2.10; CE09655.
DR InterPro: IPR001813; Ribosomal_60S.
DR InterPro: IPR001790; Ribosomal_L10.
DR Pfam: PF00428; 60S-ribosomal; 1.
DR Pfam: PF00466; Ribosomal_L10; 1.
KW Ribosomal protein; Phosphorylation.
XW INIT MET 0
SQ SEQUENCE 311 AA; 33642 MW; 2FA9A35CD24DE0F4 CRC64;

Query Match 1.6%; Score 11; DB 1; Length 311;
Best Local Similarity 100.0%; Pred. No. 0.09;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 508 AAAAPAAAAA 518
Db 275 AAAAPAAAAA 285

RESULT 4
ODP2_ALCEU STANDARD; PRT; 553 AA.
ID ODP2_ALCEU
AC Q59098;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Dihydrolipoamide acetyltransferase component of pyruvate dehydrogenase
complex (EC 2.3.1.12) [2].
GN PDHB.
OS Alkaligenes eutrophus (Ralstonia eutropha).
OC Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;
OC Burkholderiaceae; Ralstonia.
OX NCBI_TaxID=510;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=H16 / DSM 428 / ATCC 17699;
RX MEDLINE=94292470; PubMed=8021225;
RA Hein S., Steinbuechel A.;
RT "Biochemical and molecular characterization of the Alkaligenes
eutrophus pyruvate dehydrogenase complex and identification of a new
type of dihydrolipoamide dehydrogenase."
RL J. Bacteriol. 176:4394-4408(1994).
CC -!- FUNCTION: The pyruvate dehydrogenase complex catalyzes the overall
conversion of pyruvate to acetyl-CoA and CO(2). It contains
multiple copies of three enzymatic components: pyruvate
dehydrogenase (E1), dihydrolipoamide acetyltransferase (E2) and
lipoamide dehydrogenase (E3).
CC -!- CATALYTIC ACTIVITY: Acetyl-CoA + dihydrolipoamide = CoA + S-
acetyldihydrolipoamide.
CC -!- COFACTOR: Contains 2 covalently bound lipoyl cofactors (By
similarity).
CC -!- SUBUNIT: Forms a 24-polypeptide structural core with octahedral
symmetry.
CC -!- SIMILARITY: Belongs to the 2-oxoacid dehydrogenase family.
CC -!- SIMILARITY: Contains 2 lipoyl-binding domains.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).
CC

or send an email to license@isb-sib.ch).

```

C  -----
R  EMBL: U09865, AAA21599.1; --
R  PIR: B55514, B55514.
R  HSP: P10802; IDPC.
R  InterPro: IPR001078; 2Oxoacid_dh.
R  InterPro: IPR006256; Acef.
R  InterPro: IPR000089; Biotin_lipoyl.
R  InterPro: IPR003016; Lipoyl_Bs.
R  Pfam: PF00198; 2-oxoacid_dh_1.
R  Pfam: PF00384; biotin_lipoyl_2.
R  ProDom: PD00115; 2Oxoacid_dh; 1.
R  TIGRfams: TIGR01348; PDHac_trif_long; 1.
R  PROSITE: PS00189; LIPOYL; 2.
W  Glycolysis; Transferase; Acyltransferase; Repeat; Lipoyl.
T  BINDING 44 44 LIPOYL (BY SIMILARITY).
T  BINDING 162 162 LIPOYL (BY SIMILARITY).
T  ACT SITE 526 526 POTENTIAL.
T  ACT SITE 553 AA; 57338 MW; BF5D370CC60C3F12 CRC64;
Q  SEQUENCE 553 AA; 57338 MW; BF5D370CC60C3F12 CRC64;

Query Match 1.6%; Score 11; DB 1; Length 553;
Best Local Similarity 100.0%; Pred.No. 0.14;
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Y 507 AAAAAAPAAAA 517
b 199 AAAAAAPAAAA 209

RESULT 5
D NPA_PSEAM STANDARD; PRT; 82 AA.
C P04002;
T 23-OCT-1986 (Rel. 02, Created)
T 23-OCT-1986 (Rel. 02, Last sequence update)
T 16-OCT-2001 (Rel. 40, Last annotation update)
E Antifreeze protein A/B precursor.
S Pseudopleuronectes americanus (winter flounder) (Pleuronectes
S americanus).
C Eukaryota; Metazoa; Chordata; Vertebrata; Ruteleostomi;
C Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
C Acanthomorpha; Acanthopterygii; Percomorpha; Pleuronectiformes;
C Pleuronectoidae; Pleuronectidae; Pseudopleuronectes.
X NCBI_TaxID=8265;
N [1]
N SEQUENCE FROM N.A. (PROTEIN A).
P MEDLINE=82197490; PubMed=6952188;
A Davies P.L., Roach A.H., Hew C.-L.;
T "DNA sequence coding for an antifreeze protein precursor from winter
T flounder.";
L Proc. Natl. Acad. Sci. U.S.A. 79:335-339(1982).
N [2]
P SEQUENCE FROM N.A. (PROTEIN A).
X MEDLINE=88259236; PubMed=3133486;
A Scott G.K., Davies P.L., Kao M.H., Fletcher G.L.;
T "Differential amplification of antifreeze protein genes in the
T pleuronectinae.";
L J. Mol. Evol. 27:29-35(1988).
N [3]
P SEQUENCE FROM N.A. (PROTEIN B).
X MEDLINE=84264559; PubMed=6086629;
A Davies P.L., Hough C., Scott G.K., Ng N., White B.N., Hew C.-L.;
T "Antifreeze protein genes of the winter flounder.";
L J. Biol. Chem. 259:9241-9247(1984).
N [4]
P SEQUENCE FROM N.A.
X MEDLINE=92209995; PubMed=1555765;
A Davies P.L.;
T "Conservation of antifreeze protein-encoding genes in tandem
T repeats.";
L Gene 112:163-170(1992).
N [5]
P 3D-STRUCTURE MODELING OF 45-81.

```

```

RX MEDLINE=92148833; PubMed=1738160;
RT Chou K.-C.;
RT "Energy-Optimized structure of antifreeze protein and its binding
RT mechanism.";
RT J. Mol. Biol. 223:509-517(1992).
RN [6]
RX X-RAY CRYSTALLOGRAPHY (1.5 ANGSTROMS) OF 45-81.
RA Slicheri E., Yang D.S.C.;
RA MEDLINE=95281060; PubMed=7760940;
RT "Ice-binding structure and mechanism of an antifreeze protein from
RT winter flounder.";
RL Nature 375:427-431(1995).
CC -!- FUNCTION: Antifreeze proteins lower the blood freezing point.
CC -!- FUNCTION: The sequence shown is that of protein A.
CC -!- MISCELLANEOUS: BELONGS TO THE TYPB-I APP FAMILY. TYPE 1 APP ARE
CC -!- SIMILARITY: BELONGS TO THE TYPB-I APP FAMILY. TYPE 1 APP ARE
CC ALANINE-RICH, AMPHIPHILIC AND ALPHA-HELICAL.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; L00138; AAB59964.1; --
DR EMBL; L29178; AAB59964.1; JOINED.
DR EMBL; M62414; AAA49469.1; --
DR EMBL; X07506; CAA30389.1; --
DR EMBL; M62416; AAA49471.1; --
DR EMBL; M62417; AAA49472.1; --
DR PIR; A05161; A05161.
DR PIR; JS0704; FDFLAW.
DR PIR; JS0706; JS0706.
DR PIR; S02326; S02326.
DR PDB; 1ATF; 15-OCT-94.
DR PDB; 1WFA; 03-JUN-95.
DR PDB; 1WFB; 03-JUN-95.
DR InterPro; IPR000104; Antifreeze_1.
DR PRINTS; PRO0308; ANTIFREEZE1.
KW Antifreeze protein; Repeat; Multigene family; Signal; 3D-structure.
FT SIGNAL 1 21
FT PROPEP 22 44
FT CHAIN 45 82
FT VARIAT 36 36
FT VARIAT 70 70
FT CONFLICT 24 24
FT HELIX 46 80
SQ SEQUENCE 82 AA; 7711 MW; C2AE7B74C0D46CC1 CRC64;
REMOVED BY A DIPEPTIDYLPEPTIDASE
(PROBABLE).
ANTIFREEZE PROTEIN A/B.
A -> V.
A -> D (IN PROTEIN B).
S -> R (IN REF. 2).

Query Match 1.5%; Score 10; DB 1; Length 82;
Best Local Similarity 100.0%; Pred.No. 0.22;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 507 AAAAAAPAAAA 516
Db 34 AAAAAAPAAAA 43

RESULT 6
RL12_METH STANDARD; PRT; 101 AA.
AC P05394;
DT 01-NOV-1988 (Rel. 09, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE 50S ribosomal protein L12p ('A' type).
GN RPL12P OR MTH1682.
OS Methanobacterium thermoautotrophicum.
OC Archaea; Euryarchaeota; Methanobacteria; Methanobacteriales;
OC Methanobacteriaceae; Methanothermobacter.
OX NCBI_TaxID=187420;

```

RN
RP SEQUENCE FROM N.A.
RC STRAIN=Delta H;
RA MEDLINE=98037514; PubMed=9371463;
RA Smith D.R., Doucette-Stamm L.A., Deloughery C., Lee H.-M., Dubois J.,
RA Alldredge T., Bashirzadeh R., Blakely D., Cook R., Gilbert K.,
RA Harrison D., Hoang L., Keagle P., Lumm W., Pothier B., Qiu D.,
RA Spadafora R., Vicare R., Wang Y., Wierzbowski J., Gibson R.,
RA Jiwani N., Caruso A., Bush D., Safaie H., Fatwell D., Prabhakar S.,
RA McQuigall S., Shimer G., Goyal A., Pietrowski S., Church G.M.,
RA Daniels C.J., Mao J.-I., Rice P., Noelling J., Reeve J.N.;
RA "Complete genome sequence of Methanobacterium thermoautotrophicum
RA deltaH: functional analysis and comparative genomics.";
RL J. Bacteriol. 179:7135-7155 (1997).
[2]
RN SEQUENCE OF 1-48.
RP MEDLINE=8110590; PubMed=706702;
RA Matheson A.T., Yaguchi M., Balch W.B., Wolfe R.S.;
RA "Sequence homologs in the N-terminal region of the ribosomal 'A'
RA proteins from Methanobacterium thermoautotrophicum and Halobacterium
RA cutirubrum";
RL Biochim. Biophys. Acta 626:162-169 (1980).
RC -!- FUNCTION: Seems to be the binding site for several of the factors
RC involved in protein synthesis and appears to be essential for
RC accurate translation (By similarity).
CC -!- SIMILARITY: Belongs to the L12P family of ribosomal proteins.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

CC EMBL: A3000926; AAB86154.1; -
CC PIR: H69091; H69091.
CC InterPro: IPR001813; Ribosomal_60S.
CC Pfam: PF00428; 60s_ribosomal; 1.
CC Ribosomal protein; Complete proteome.
CC SEQUENCE 101 AA; 10480 MW; 6C1CEA3BF121FB85 CRC64;

Query Match 1.5%; Score 10; DB 1; Length 101;
Best Local Similarity 100.0%; Pred. No. 0.26; Mismatches 0; Indels 0; Gaps 0;
Matches 10; Conservative 0;
Y 509 AAAAAAAAAA 518
| | | | | | | | | |
b 60 AAAAAAAAAA 69

RESULT 7
RL12 METTL
ID RL12 METTL STANDARD; PRT; 101 AA.
AC 052706;
RT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 30-MAY-2004 (Rel. 39, Last annotation update)
DE 50S ribosomal protein L12P.
RN Methanococcus thermolithotrophicus.
RC Archaea; Buryarchaeota; Methanococci; Methanococcales;
CC Methanococcaceae; Methanothermococcus.
CC NCBI_TaxID=2186;
[1]
RN SEQUENCE FROM N.A.
RA Linhart A., Fiendi W.;
RA Submitted (JUN-1998) to the EMBL/GenBank/DBJ databases.
RC -!- FUNCTION: Seems to be the binding site for several of the factors
RC involved in protein synthesis and appears to be essential for
RC accurate translation (By similarity).
CC -!- SIMILARITY: Belongs to the L12P family of ribosomal proteins.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

CC EMBL: AF044919; AAC64512.1; -
CC InterPro: IPR001813; Ribosomal_60S.
CC Pfam: PF00428; 60s_ribosomal; 1.
CC Ribosomal protein.
CC SEQUENCE 101 AA; 10292 MW; 27262815729C8CB0 CRC64;

Query Match 1.5%; Score 10; DB 1; Length 101;
Best Local Similarity 100.0%; Pred. No. 0.26; Mismatches 0; Indels 0; Gaps 0;
Matches 10; Conservative 0;
Y 506 VAAAAAPAAA 515
| | | | | | | | | |
b 61 VAAAAAPAAA 70

RESULT 8
RLA2 DICDI
ID RLA2 DICDI STANDARD; PRT; 105 AA.
AC P22683;
DT 01-AUG-1991 (Rel. 19, Created)
DT 01-NOV-1991 (Rel. 20, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE 60S acidic ribosomal protein P2.
RN Dictyostelium discoideum (Slime mold).
RC Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
CC NCBI_TaxID=44689;
[1]
RN SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RC STRAIN=AX3;
RX MEDLINE=91232921; PubMed=1840653;
RA Prieto J., Candel E., Coloma A.;
RT "Nucleotide sequence of a cDNA encoding acidic ribosomal
RL phosphoprotein P2 in Dictyostelium discoideum";
RL Nucleic Acids Res. 19:1341-1341(1991).
CC -!- FUNCTION: Plays an important role in the elongation step of
CC protein synthesis.
CC -!- SUBUNIT: P1 and P2 exist as dimers at the large ribosomal subunit.
CC -!- PTM: Phosphorylated.
CC -!- SIMILARITY: Belongs to the L12P family of ribosomal proteins.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

CC EMBL: X56192; CAA39655.1; -
CC PIR: S14014; R6DOP2
CC DictyBase: DB0001822; X56192.
CC InterPro: IPR001813; Ribosomal_60S.
CC Pfam: PF00428; 60s_ribosomal; 1.
CC Ribosomal protein; Phosphorylation.
CC INIT MET 0
CC SEQUENCE 105 AA; 10378 MW; 519FAB6679A5B840 CRC64;

Query Match 1.5%; Score 10; DB 1; Length 105;
Best Local Similarity 100.0%; Pred. No. 0.27; Mismatches 0; Indels 0; Gaps 0;
Matches 10; Conservative 0;
Y 506 VAAAAAPAAA 515
| | | | | | | | | |
b 62 VAAAAAPAAA 71

```

GN RPLL OR LMO0251.
OS Listeria monocytogenes.
OC Bacteria; Firmicutes; Bacillales; Listeriaceae; Listeria.
OX NCBI_TaxID=1639;
RN (1)
RP SEQUENCE FROM N.A.
RC STRAIN=EGD-e / Serovar 1/2a;
RX MEDLINE=21537279; PubMed=11679669;
RA Glaser P., Frangeul L., Buchrieser C., Rusniok C., Amend A.,
RA Baquero P., Berche P., Bloeker H., Brandt P., Chakraborty T.,
RA Charbit A., Chetouani F., Couve E., de Daruvar A., Deboux P.,
RA Domann E., Dominguez-Bernal G., Duchaud E., Durant L., Dussurget O.,
RA Entian K.-D., Fshih H., Garcia-del Portillo F., Garrido P.,
RA Gautier L., Goebel W., Gomez-Lopez N., Hain T., Hauf J., Jackson D.,
RA Jones L.-M., Kaerst U., Kreft J., Kuhn M., Kunst F., Kurapkat G.,
RA Madueno E., Maitournam A., Mata Vicente J., Ng E., Nedjari H.,
RA Nordsiek G., Novella S., de Pablo B., Perez-Diaz J.-C., Purcell R.,
RA Remmel B., Rose M., Schlueter T., Simoes N., Tierrez A.,
RA Vazquez-Boland J.-A., Voss H., Wehland J., Cossart P.,
RT "Comparative genomics of Listeria species.";
RL Science 294:849-852(2001).
CC -!- FUNCTION: Seems to be the binding site for several of the factors
CC involved in protein synthesis and appears to be essential for
CC accurate translation (By similarity).
CC -!- SIMILARITY: Belongs to the L12p family of ribosomal proteins.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AL591974; CAD00778.1; -
CC PIR; AD1106; AD1106.
CC ListList; LMO0251; -
CC HAMAP; MF_00368; -; 1.
CC InterPro; IPR008932; Ribos_L12/7_olig.
CC InterPro; IPR00206; Ribosomal_L12.
CC Pfam; PF00542; Ribosomal_L12; 1.
CC ProDom; PD001326; Ribosomal_L12; 1.
CC TIGRFAMs; TIGR00855; L12; 1.
CC Ribosomal protein; Complete proteome.
CC SEQUENCE 120 AA; 12469 MW; 8BC3D5417F6B1954 CRC64;

Query Match 1.5%; Score 10; DB 1; Length 120;
Best Local Similarity 100.0%; Pred.No. 0.3;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 500 AAAPVAVAAA 509
DB 35 AAAPVAVAAA 44

RESULT 11
RL7 STRPN
ID RL7 STRPN STANDARD; PRT; 121 AA.
AC P80714;
DT 01-OCT-1996 (Rel. 34, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE 50S ribosomal protein L7/L12.
GN RPLL OR SPL354 OR SPL211.
OS Streptococcus pneumoniae, and
OS Streptococcus pneumoniae (strain ATCC BAA-255 / R6).
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1313, 171101;
RN (1)
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC BAA-334 / TIGR4;
RX MEDLINE=21357209; PubMed=11463916;

```

RA Tettelin H., Nelson K.E., Paulsen I.T., Eisen J.A., Read T.D.,
 RA Peterson S., Heidelberg J., DeBoy R.T., Haft D.H., Dodson R.J.,
 RA Durkin A.S., Whitte M., Kolonay J.F., Nelson W.C., Peterson J.D.,
 RA Umayam L.A., White O., Salanbay S.L., Lewis M.R., Radune D.,
 RA Holtzapple E., Khouri H., Wolf A.M., Utterback T.R., Hansen C.L.,
 RA McDonald L.A., Feldblyum T.V., Angiuoli S., Dickinson T., Hickey B.K.,
 RA Holt I.E., Loftus B.J., Yang P., Smith H.O., Venter J.C., Hickey B.K.,
 RA Dougherty B.A., Morrison D.A., Hollingshead S.K., Fraser C.M.;
 RA "Complete genome sequence of a virulent isolate of Streptococcus
 pneumoniae.";
 RT Science 293:498-506(2001).
 RL [2]
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=ATCC BAA-255 / R6;
 RX MEDLINE=21429245; PubMed=11544234;
 RA Hoskins J., Aborn W.E. Jr., Arnold J., Blaszcak L.C., Burgett S.,
 RA DeHoff B.S., Estrem S.T., Fritz L., Fu D.-J., Fuller W., Geringer C.,
 RA Gilmore R., Glass J.S., Khoja H., Kraft A.R., Lagace R.E.,
 RA LeBlanc D.J., Lee L.N., Lefkowitz E.J., Lu J., Matsushima P.,
 RA McAbren S.M., McHenry M., McLeaster K., Mundy C.W., Niclas T.I.,
 RA Norris F.H., O'Gara M., Peary R.B., Robertson G.T., Rockey P.,
 RA Sun P.-M., Winkler M.E., Yang Y., Young-Bellido M., Zhao G.,
 RA Zook C.A., Baltz R.H., Jaskunas S.R., Rostock P.R. Jr., Skatrud P.L.,
 RA Glass J.I.;
 RA "Genome of the bacterium Streptococcus pneumoniae strain R6.";
 RT J. Bacteriol. 183:5709-5717(2001).
 RL [3]
 RN [3]
 RP SEQUENCE OF 1-42.
 RC MEDLINE=9717772; PubMed=9025278;
 RX Kolberg J., Heiby E.A., Lopez R., Sletten K.;
 RA "Monoclonal antibodies against Streptococcus pneumoniae detect
 RT epitopes on eubacterial ribosomal proteins L7/L12 and on
 RT streptococcal elongation factor Ts.";
 RL Microbiology 143:55-61(1997).
 CC -!- FUNCTION: Seems to be the binding site for several of the factors
 CC involved in protein synthesis and appears to be essential for
 CC accurate translation [By similarity].
 CC -!- SIMILARITY: Belongs to the L12P family of ribosomal proteins.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announcement/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL: AE007433; AAK75452.1; -;
 CC EMBL: AE008493; AAL00015.1; ALT_INIT.
 CC PIR: C95157; C95157.
 CC TIGR: SP1354;
 CC HAMAP: MF_00368; -; 1.
 CC InterPro: IPR008932; Ribos_L12/7_olig.
 CC InterPro: IPR00206; Ribosomal_L12.
 CC Pfam: PF00542; Ribosomal_L12; 1.
 CC ProDom: PD001326; Ribosomal_L12; 1.
 CC TIGRFAMs: TIGR00855; L12; 1_
 CC Ribosomal protein; Complete proteome.
 CC INIT_MET 0
 CC SEQUENCE 121 AA; 12311 MW; E284993FB326E0D0 CRC64;
 SQ
 Query Match 1.5%; Score 10; DB 1; Length 121;
 Best Local Similarity 100.0%; Pred.No. 0.3;
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 500 AAAPVAVAAA 509
 DB 34 AAAPVAVAAA 43
 RESULT 12
 RL7_FUSNN
 ID RL7_FUSNN STANDARD; PRT; 122 AA.

AC Q8RH15;
 DT 28-FEB-2003 (Rel. 41, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE 50S ribosomal protein L7/L12.
 GN RELL OR FN2037.
 OS Fusobacterium nucleatum (subsp. nucleatum).
 OC Bacteria; Fusobacteria; Fusobacteriales; Fusobacteriaceae;
 OC Fusobacterium.
 OX NCBI_TaxID=76856;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=ATCC 25586;
 RX MEDLINE=21886394; PubMed=11889109;
 RA Kapral V., Anderson I., Ivanova N., Reznik G., Los T., Lykidis A.,
 RA Bhattacharya A., Bartman A., Gardner W., Grechkin G., Zhu L.,
 RA Vasileva O., Chu L., Kogan Y., Chaga O., Goltsman E., Bernal A.,
 RA Larsen N., D'Souza M., Walunas T., Fusch G., Haselkorn R.,
 RA Fongstein M., Kyrpides N., Overbeek R.;
 RA "Genome sequence and analysis of the oral bacterium Fusobacterium
 RT nucleatum strain ATCC 25586.";
 RL J. Bacteriol. 184:2005-2018(2002).
 CC -!- FUNCTION: Seems to be the binding site for several of the factors
 CC involved in protein synthesis and appears to be essential for
 CC accurate translation [By similarity].
 CC -!- SIMILARITY: Belongs to the L12P family of ribosomal proteins.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announcement/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL: AE010507; AAL94122.1; -;
 CC HAMAP: MF_00368; -; 1.
 CC InterPro: IPR008932; Ribos_L12/7_olig.
 CC InterPro: IPR00206; Ribosomal_L12.
 CC Pfam: PF00542; Ribosomal_L12; 1.
 CC ProDom: PD001326; Ribosomal_L12; 1.
 CC TIGRFAMs: TIGR00855; L12; 1_
 CC Ribosomal protein; Complete proteome.
 CC SEQUENCE 122 AA; 12605 MW; 4382328A2B266BA3 CRC64;
 SQ
 Query Match 1.5%; Score 10; DB 1; Length 122;
 Best Local Similarity 100.0%; Pred.No. 0.31;
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 500 AAAPVAVAAA 509
 DB 35 AAAPVAVAAA 44
 RESULT 13
 RL7_BRUME
 ID RL7_BRUME STANDARD; PRT; 124 AA.
 AC P41106; Q9R2F1;
 DT 01-FEB-1995 (Rel. 31, Created)
 DT 01-FEB-1995 (Rel. 31, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE 50S ribosomal protein L7/L12.
 GN RPLL OR BME10748 OR BR1245.
 OS Brucella melitensis,
 OS Brucella suis, and
 OS Brucella abortus.
 OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
 OC Brucellaceae; Brucella.
 OX NCBI_TaxID=29459, 29461, 235;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC SPECIES=B.melitensis; STRAIN=16M / ATCC 23456 / Biotype 1;
 RX MEDLINE=94357413; PubMed=8076798;

BAchrach G., Bar-Nir D., Banai M., Bercovier H.;
"Identification and nucleotide sequence of Brucella melitensis L7/L12
ribosomal protein";
FEMS Microbiol. Lett. 120:237-240(1994).
[2]
SEQUENCE FROM N.A.
SPECIES=B.melitensis; STRAIN=16M / ATCC 23456 / Biotype 1;
MEDLINE=20020109; PubMed=11756688;
A Delvecchio V.G., Kapural V., Redkar R.J., Patra G., Muej C., Los T.,
A Ivanova N., Anderson I., Bhattacharyya A., Lykidis A., Reznik G.,
A Jablonski L., Larsen N., D'Souza M., Bernal A., Mazur M., Goldsman E.,
A Selkov E., Elzer P.H., Hagius S., O'Callaghan D., Letesson J.-J.,
A Haselkorn R., Kypides N., Overbeek R.;
"The genome sequence of the facultative intracellular pathogen
Brucella melitensis";
Proc. Natl. Acad. Sci. U.S.A. 99:443-448(2002).
[3]
SEQUENCE FROM N.A.
SPECIES=B.suis; STRAIN=1330 / Biovar 1;
MEDLINE=2247741; PubMed=12271122;
A Paulsen I.T., Seshadri R., Nelson K.E., Eisen J.A., Heidelberg J.F.,
A Read T.D., Dodson R.J., Umayam L., Brinkac L.M., Beanan M.J.,
A Daugherty S.C., Deboy R.T., Durkin A.S., Kolonay J.F., Madupu R.,
A Nelson W.C., Ayodeji B., Kraul M., Shetty J., Malek J., Van Aken S.E.,
A Redmiller S., Tettelin H., Gill S.R., White O., Salzberg S.B.,
A Hoover D.L., Lindler L.E., Hailing S.M., Boyle S.M., Fraser C.M.;
"The Brucella suis genome reveals fundamental similarities between
animal and plant pathogens and symbionts";
Proc. Natl. Acad. Sci. U.S.A. 99:13148-13153(2002).
[4]
SEQUENCE FROM N.A.
SPECIES=B.abortus; STRAIN=19;
MEDLINE=94171071; PubMed=812531;
A Oliveira S.C., Zhu Y., Splitter G.A.;
"Sequences of the rplJL operon containing the L10 and L7/L12 genes
from Brucella abortus";
Gene 140:137-138(1994).
[5]
SEQUENCE FROM N.A.
SPECIES=B.abortus; STRAIN=19;
A Campos E., Cravero S., Boschirol M.L., Arese A.I., Rossetti O.L.;
"Genetic organization of the rplJL-rpoB operon in Brucella abortus";
(In) Frank J.P. (eds.);
Networking in Brucellosis Research II - Proceedings of the UNU-BIOLAC
Brucellosis Workshop, pp.1-8, United Nations University Press, Tokyo
(1998).
-I- FUNCTION: Seems to be the binding site for several of the factors
involved in protein synthesis and appears to be essential for
accurate translation (By similarity).
-I- SIMILARITY: Belongs to the L12P family of ribosomal proteins.

This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).

EMBL; L27819; AAA56790.1; --
EMBL; A0009516; AAL51929.1; --
EMBL; L19101; AAL19863.1; --
EMBL; AF169147; AAD51621.1; --
EMBL; A5014423; AAN30164.1; --
PIR; AF3345; AF3345;
PIR; I40348; I40348;
PIR; I40350; I40350;
HSSP; P02392; ICTP.
TIGR; BR1245; --
HAMAP; MF_00368; -- 1;
InterPro; IPR008932; Ribos L12/7 olig.
InterPro; IPR00206; Ribosomal L12.
Pfam; PF00542; Ribosomal_L12; 1.

DR ProDom; PD001326; Ribosomal_L12; 1.
DR TIGRPFMS; TIGR00855; L12; 1.
RW Ribosomal protein; Complete proteome.
FT CONFLICT 104 K -> N (in REF. 5).
SQ SEQUENCE 124 AA; 12546 MW; A56E1079AABBE7F4 CRC64;

Query Match 1.5%; Score 10; DB 1; Length 124;
Best Local Similarity 100.0%; Pred. NO. 0.31;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 500 AAPAVAVAAA 509
Db 34 AAPAVAVAAA 43
|||||||
|||

RESULT 14
RL7_AGR75
ID - RL7_AGR75 STANDARD; PRT; 125 AA.
AC Q8UE07;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE 50S ribosomal protein L7/L12.
GN RPLL OR ARJ1957 OR AGH_C_3571.
OS Agrobacterium tumefaciens (strain C58 / ATCC 33970).
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
OC Rhizobiaceae; Rhizobium; Agrobacterium group; Agrobacterium.
OX NCBI TaxID=176299;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21608550; PubMed=11743193;
RA Wood D.W., Setubal J.C., Kaul R., Monks D.E., Kitajima J.P.,
RA Okura Y., Zhou Y., Chen L., Wood G.B., Almeida N.F. Jr., Woo L.,
RA Chen Y., Paulsen I.T., Eisen J.A., Karp P.D., Bovee D. Sr.,
RA Chapman P., Clendenning J., Deatherage G., Gillet W., Grant C.,
RA Kutayavin T., Levy R., Li M.-J., McClelland E., Palmieri A.,
RA Raymond C., Rouse G., Saenphimmachak C., Wu Z., Romero P., Gordon D.,
RA Zhang S., Yoo H., Tao Y., Biddle P., Jung M., Krespan W., Perry M.,
RA Gordon-Kamm B., Liao L., Kim S., Hendrick C., Zhao Z.-Y., Dolan M.,
RA Chumley F., Tingey S.V., Tomb J.-F., Gordon M.P., Olson M.V.,
RA Nester E.W.;
"The genome of the natural genetic engineer Agrobacterium tumefaciens
C58";
Science 294:2317-2323(2001).
[2]
RP SEQUENCE FROM N.A.
RX MEDLINE=21608551; PubMed=11743194;
RA Goodner B., Hinkle G., Gattung S., Miller N., Blanchard M.,
RA Ourlo B., Goldman B.S., Cao Y., Askenazi M., Halling C., Mullin L.,
RA Houmuel K., Gordon J., Vaudin M., Iartchouk O., Spp A., Liu P.,
RA Wollam C., Allinger M., Doughty D., Scott C., Lappas C., Markelz B.,
RA Flanagan C., Crowell C., Gurson J., Lomo C., Sear C., Strub G.,
RA Cielo C., Slater S.;
"Genome sequence of the plant pathogen and biotechnology agent
Agrobacterium tumefaciens C58";
Science 294:2323-2328(2001).
RL Science 294:2323-2328(2001).
CC -I- FUNCTION: Seems to be the binding site for several of the factors
involved in protein synthesis and appears to be essential for
accurate translation (By similarity).
CC -I- SIMILARITY: Belongs to the L12P family of ribosomal proteins.

This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).

EMBL; A0009149; AAL42953.1; --
EMBL; A0008114; AAK87717.1; --
PIR; AC2817; AC2817.
PIR; D97595; D97595.

Search completed: June 16, 2004, 19:26:11
Job time : 18 secs

```
DR HAMAP: MF_00368; -: 1.
DR InterPro: IPR008932; Ribos_L12/7 olig.
DR InterPro: IPR000206; Ribosomal_L12.
DR Pfam: PF00542; Ribosomal_L12; 1.
DR ProDom: PD001326; Ribosomal_L12; 1.
DR TIGRFAMs: TIGR00855; L12; 1.
KW Ribosomal protein; Complete proteome.
SQ SEQUENCE 125 AA; 12736 MW; 37A8E02FE77175D CRC64;

Query Match 1.58; Score 10; DB 1; Length 125;
Best Local Similarity 100.0%; Pred. NO. 0.31;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

2Y 500 AAAPVAVAAA 509
2b 34 AAAPVAVAAA 43

RESULT 15
2L7_RHILO STANDARD; PRT; 125 AA.
AC Q98N67;
YT 28-FEB-2003 (Rel. 41, Created)
YT 28-FEB-2003 (Rel. 41, Last sequence update)
YT 28-FEB-2003 (Rel. 41, Last annotation update)
2E 50S ribosomal protein L7/L12.
2N RPLL OR MLR0275.
2S Rhizobium loti (Mesorhizobium loti).
2C Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
2X Phyllobacteriaceae; Mesorhizobium.
2X NCBI_TaxID=381;
2N (1)
2P SEQUENCE FROM N.A.
2X STRAIN=MAFF303099;
2X MEDLINE=21082930; PubMed=11214968;
2A Kaneko T., Nakamura Y., Sato S., Asamizu E., Kato T., Sasamoto S.,
2A Watanabe A., Idegawa K., Ishikawa A., Kawashima K., Kimura T.,
2A Kishida Y., Kiyokawa C., Kohara M., Matsumoto M., Matsuno A.,
2A Mochizuki Y., Nakayama S., Nakazaki N., Shimpo S., Sugimoto M.,
2A Takeuchi C., Yamada M., Tabata S.;
2X "Complete genome structure of the nitrogen-fixing symbiotic bacterium
2T Mesorhizobium loti."
2L DNA Res. 7:331-338(2000).
2C -!- FUNCTION: Seems to be the binding site for several of the factors
2C involved in protein synthesis and appears to be essential for
2C accurate translation (by similarity).
2C -!- SIMILARITY: Belongs to the L12P family of ribosomal proteins.
2C -----
2C This SWISS-PROT entry is copyright. It is produced through a collaboration
2C between the Swiss Institute of Bioinformatics and the EMBL outstation -
2C the European Bioinformatics Institute. There are no restrictions on its
2C use by non-profit institutions as long as its content is in no way
2C modified and this statement is not removed. Usage by and for commercial
2C entities requires a license agreement (see http://www.isb-sib.ch/announce/
2C or send an email to license@isb-sib.ch).
2C -----
2C EMBL: AP002994; BAB47895.1; -.
2R HAMAP: MF_00368; -: 1.
2R InterPro: IPR008932; Ribos_L12/7 olig.
2R InterPro: IPR000206; Ribosomal_L12.
2R Pfam: PF00542; Ribosomal_L12; 1.
2R ProDom: PD001326; Ribosomal_L12; 1.
2R TIGRFAMs: TIGR00855; L12; 1.
KW Ribosomal protein; Complete proteome.
SQ SEQUENCE 125 AA; 12735 MW; 9BE14454212A55AF CRC64;

Query Match 1.58; Score 10; DB 1; Length 125;
Best Local Similarity 100.0%; Pred. NO. 0.31;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

2Y 500 AAAPVAVAAA 509
2b 34 AAAPVAVAAA 43
```


CC be used to manufacture medicaments for treating angiogenesis-related
CC diseases or disorders, such as tumor conditions, diabetes, rheumatoid
CC arthritis, and even some inflammatory diseases such as psoriasis, chronic
CC inflammation of the intestine, asthma, etc. . The protein may also be able
CC to treat and cure, or prevent, obesity. The ABP-1 DNA can be used in gene
CC therapy techniques
XX
SQ Sequence 675 AA;

Query Match 100.0%; Score 675; DB 3; Length 675;
Best Local Similarity 100.0%; Pred. NO. 0;
Matches 675; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

2Y 1 MPRAQSSASVOPADPFAIVSAQOMVILSDENNLFOELGCGVKEVARLQKVETRI 60
2b 1 MPRAQSSASVOPADPFAIVSAQOMVILSDENNLFOELGCGVKEVARLQKVETRI 50
2Y 61 QRVSEAVENLVKSSKREALEKAMRNKLEGEIRRMHDFNRDLERLETANKQLAEKEYEG 120
2b 61 QRVSEAVENLVKSSKREALEKAMRNKLEGEIRRMHDFNRDLERLETANKQLAEKEYEG 120
2Y 121 SEDTRKTIISQIFAKNKESOREKEKLEALATARSTNEDQRRHLEIRDOALSNAQAKVKL 180
2b 121 SEDTRKTIISQIFAKNKESOREKEKLEALATARSTNEDQRRHLEIRDOALSNAQAKVKL 180
2Y 181 EELKCKQVYVDKVKRQQAALVOLQAACEKEQLHRLTRLERELSLRIQORQNCQP 240
2b 181 EELKCKQVYVDKVKRQQAALVOLQAACEKEQLHRLTRLERELSLRIQORQNCQP 240
2Y 241 TNVSEYNAALMELLREKEERILALEADMTKWEQKYLEENVMHFALDAATVAARDTT 300
2b 241 TNVSEYNAALMELLREKEERILALEADMTKWEQKYLEENVMHFALDAATVAARDTT 300
2Y 301 VISHSPNTSYDTALEARIOKEEERILMANKECLDMGRIKTLHAQIIEKDAMIKVLQORS 360
2b 301 VISHSPNTSYDTALEARIOKEEERILMANKECLDMGRIKTLHAQIIEKDAMIKVLQORS 360
2Y 361 RKEPSKTQLSCMPAPKSLMSISNAGSGLLSHSSTLTGSPIMEKRDKSKWSGLILIG 420
2b 361 RKEPSKTQLSCMPAPKSLMSISNAGSGLLSHSSTLTGSPIMEKRDKSKWSGLILIG 420
2Y 421 GDYRAEYVSPSPVPPSTPLLSAHSKGTGSDCSTQTERGTESNKTAAVAPISVPAPVAA 480
2b 421 GDYRAEYVSPSPVPPSTPLLSAHSKGTGSDCSTQTERGTESNKTAAVAPISVPAPVAA 480
2Y 481 AATAAATATAATITTTWVAAPVAVAAAAAPAAAAAPSPATAATAAASPAAAGQIPA 540
2b 481 AATAAATATAATITTTWVAAPVAVAAAAAPAAAAAPSPATAATAAASPAAAGQIPA 540
2Y 541 AASVASAAVAPSAASAAAQVAPAPVAPALVPVAPAPAAQAQAPACTQAPTSA 600
2b 541 AASVASAAVAPSAASAAAQVAPAPVAPALVPVAPAPAAQAQAPACTQAPTSA 600
2Y 601 VAPTPTPTTFAVAQAEVAPSPATGPGPHRLSIPSLTCNPKDGTGPFVHNTLERTPIQ 660
2b 601 VAPTPTPTTFAVAQAEVAPSPATGPGPHRLSIPSLTCNPKDGTGPFVHNTLERTPIQ 660
2Y 661 ILGQEPDAEWVEYLI 675
2b 661 ILGQEPDAEWVEYLI 675

RESULT 2
ID AAY54053
XX AAY54053 standard; protein; 675 AA.
AC AAY54053;
XX
XX
XX 27-MAR-2000 (first entry)
XX A variant of an angiogenesis-associated protein which binds plasminogen.
DE
XX Human; angiogenesis-associated protein; plasminogen; ABP-1;

KW kringle domain; angiotatin; plasminogen receptor;
KW angiogenesis-related disease; tumor; diabetes; rheumatoid arthritis;
KW inflammatory disease; psoriasis; chronic inflammation; intestine; asthma;
XX obesity; gene therapy.
OS Homo sapiens.
XX
XX Key Location/Qualifiers
FT Misc-difference 135
FT Misc-difference /label= Asn, Ser, Asp
FT Misc-difference 148. .150
FT /note= "these residues are either Glu-Leu-Ala or Thr-Thr-Pro"
XX
XX WO9966038-A1.
XX 23-DEC-1999.
XX 11-JUN-1999; 99WO-BP004109.
XX 15-JUN-1998; 98SE-00002130.
XX 15-JUN-1998; 98US-0089266P.
XX 17-DEC-1998; 98SE-00004372.
XX 29-DEC-1998; 98US-0114386P.
XX (PHAA) PHARMACIA & UPJOHN AB.
XX Holmgren L, Troyanovsky B;
XX WPI; 2000-106099/09.
XX Novel human protein useful for treating angiogenesis associated diseases or disorders.
XX Claim 5; Page 46-49; 58pp; English.
XX The present sequence represents a polymorphic variant of a human
XX angiogenesis-associated protein which is able to bind an N-terminal
XX fragment of plasminogen. The native protein, described in AAY54052 is
XX designated ABP-1, and binds the first 4 kringle domains (K1-K4) and/or
XX kringle 5 (K5) of plasminogen. These four kringle domains comprise
XX angiotatin-binding domain of the ABP-1 protein is described in AAY54054.
XX angiotatin-binding domain of the ABP-1 protein is described in AAY54054.
XX ABP-1 can be used to manufacture medicaments for treating angiogenesis-
XX related diseases or disorders, such as tumor conditions, diabetes,
XX rheumatoid arthritis, and even some inflammatory diseases such as
XX psoriasis, chronic inflammation of the intestine, asthma, etc. . The
XX protein may also be able to treat and cure, or prevent, obesity. The ABP-
XX 1 DNA can be used in gene therapy techniques

SQ Sequence 675 AA;
Query Match 77.8%; Score 525; DB 3; Length 675;
Best Local Similarity 100.0%; Pred. NO. 0;
Matches 525; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 151 TARSTNEQRRHIEIRDOALSNAQAKVKLEELKQVYVDKVKRQQAALVOLQAACEK 210
DB 151 TARSTNEQRRHIEIRDOALSNAQAKVKLEELKQVYVDKVKRQQAALVOLQAACEK 210
QY 211 REQLHRLTRLERELSLRIQORQNCQPTNVSEYNAAALMELLREKEERILAEADMT 270
DB 211 REQLHRLTRLERELSLRIQORQNCQPTNVSEYNAAALMELLREKEERILAEADMT 270
QY 271 KWEQKYLEENVMHFALDAATVAARDTTVISHSPNTSYDTALEARIOKEEERILMAN 330
DB 271 KWEQKYLEENVMHFALDAATVAARDTTVISHSPNTSYDTALEARIOKEEERILMAN 330
QY 331 RCLDMGRIKTLHAQIIEKDAMIKVLQORSKESKTEQLSCMPAPKSLMSISNAGSGL 390
DB 331 RCLDMGRIKTLHAQIIEKDAMIKVLQORSKESKTEQLSCMPAPKSLMSISNAGSGL 390
QY 391 SHSSTLTGSPIMEKRDKSKWSGLILIGDYAEVVPSTPSPVPPSTPLLSAHSKGTGS 450

```

391 SHSSTLGTSPIMEXKDDKSKGSLGILLGGDYAEVPSPTSPVPPSTLLSAHSKTS 450
451 RDCSTQTERGTESKTAAPVAPISVPAPVAAAATAAATTAATITTTTMAAPVAAVAAA 510
451 RDCSTQTERGTESKTAAPVAPISVPAPVAAAATAAATTAATITTTTMAAPVAAVAAA 510
511 AAAAAAPSPATAAATAAASPAAGQIPAAASVASAAAVASAAAAVQVAPAPV 570
511 AAAAAAPSPATAAATAAASPAAGQIPAAASVASAAAVASAAAAVQVAPAPV 570
571 PAPALVVPAPAAAQASAPAQTAQTSAPAVATPPTPTTTPAVAQAEVSPATPGPHR 630
571 PAPALVVPAPAAAQASAPAQTAQTSAPAVATPPTPTTTPAVAQAEVSPATPGPHR 630
631 LSIPSLTCNPKDKTGPVPHSNTLERTPIQILQGBDPAEMVEYLI 675
631 LSIPSLTCNPKDKTGPVPHSNTLERTPIQILQGBDPAEMVEYLI 675

```

```

RESULT 3
AAE37918 standard; protein; 993 AA.
AAE37918;
06-NOV-2003 (first entry)
Human CGDD-7 protein.
Human; cell growth, differentiation and death protein; CGDD; leukaemia;
neurodegenerative disorder; Parkinson's disease; Alzheimer's disease;
muscular disorder; myotonic dystrophy; catatonia; endocrine disorder;
diabetes; Grave's disease; cancer; immunological disorder; scleroderma;
systemic lupus erythematosus; allergy; Crohn's disease; renal disorder;
gastrointestinal disorder; Goodpasture's syndrome; infection; cirrhosis;
cardiovascular disorder; atherosclerosis; hepatic disease; transgenic;
transgenic animal; gene therapy; neuroprotective; relaxant; cytostatic;
dermatological; immunosuppressive; cerebroprotective; anticonvulsant;
antibacterial; antiparasitic; fungicide; virucide; uropathic; cardiant;
protozoacide; nootropic.
Homo sapiens.
WO2003050253-A2.
19-JUN-2003.
04-DEC-2002; 2002WO-US039133.
07-DEC-2001; 2001US-0340747P.
20-DEC-2001; 2001US-0342761P.
15-JAN-2002; 2002US-0349705P.
06-FEB-2002; 2002US-0354764P.
12-FEB-2002; 2002US-0356216P.
(INCY-) INCYTE GENOMICS INC.
Griffin JA, Runkumar J, Emerling BM, Kable AE, Elliott VS,
Marquis JP, Baughn MR, Corvad AE, Yue H, Lee EA, Becha SD, Tang YT,
Tran UK, Swarnakar A, Lee S, Ison CH, Hafalia AJA, Tran B,
Sprague WW, Lee SY, Khare R, Gandhi AR, Gietzen KJ, Bhatia U,
Burrill JD, Blake JJ, Ho A, Zheng W,
WPI; 2003-532903/50.
N-PSDB; AAD57227.
New CGDD polypeptides, useful for diagnosing, preventing, and treating
disorders associated with an abnormal expression or activity of CGDD,
e.g. neuromuscular, immunological, cardiovascular disorders, cancer
and/or infections.
Claim 1; Page 220-222; 299pp; English.

```

```

XX The present invention relates to novel cell growth, differentiation and
CC death (CGDD) proteins and polynucleotides encoding them. The sequences of
CC the invention are useful in diagnosing, preventing and treating disorders
CC associated with an abnormal expression or activity of CGDD such as
CC neurodegenerative disorders (e.g. Parkinson's disease, Alzheimer's
CC disease), muscular disorders (e.g. myotonic dystrophy, catatonia),
CC endocrine disorders (e.g. diabetes, Grave's disease), cancers (e.g.
CC leukaemia, cervical or breast cancers), immunological disorders (e.g.
CC scleroderma, systemic lupus erythematosus, allergies), gastrointestinal
CC disorders (e.g. Crohn's disease), renal disorders (e.g. Goodpasture's
CC syndrome), infections (e.g. viral, bacterial, fungal, parasitic,
CC protozoal, helminthic), cardiovascular disorders (e.g. atherosclerosis)
CC and hepatic diseases (e.g. cirrhosis). The polynucleotides can be used to
CC create humanised animals or transgenic animals to model human diseases.
CC The invention is also used in gene therapy. The present sequence is human
CC CGDD-7 protein
XX Sequence 993 AA;
SQ
Query Match 72.9%; Score 492; DB 7; Length 993;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 492; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MPRAQSSASQYVPADPFAIVSRAQQMVILSDENRNLROELGECYKVARLOKVETEI 60
DB 319 MPRAQSSASQYVPADPFAIVSRAQQMVILSDENRNLROELGECYKVARLOKVETEI 378
QY 61 QRYSEAVENLVKSSKREALEKAMRNKLEGEIRRMHDFNRDLRERLETANKQLAEKEYEG 120
DB 379 QRYSEAVENLVKSSKREALEKAMRNKLEGEIRRMHDFNRDLRERLETANKQLAEKEYEG 438
QY 121 SEDTRKTIISQIFANKKESQREKLEAEATARSNTDQRRHIEIRDQALSNAQKVVKL 180
DB 439 SEDTRKTIISQIFANKKESQREKLEAEATARSNTDQRRHIEIRDQALSNAQKVVKL 498
QY 181 EERUKKQVTVVKVMQOALVOLQACEKQEOLERLRLTRERLESRLIQORQNCQP 240
DB 499 EELKXQVTVVKVMQOALVOLQACEKQEOLERLRLTRERLESRLIQORQNCQP 558
QY 241 TNVSEYNAALMBELLREKERILALEADMTKWQKYLEBNVMHFPALDAAATVAAQRDTT 300
DB 559 TNVSEYNAALMBELLREKERILALEADMTKWQKYLEBNVMHFPALDAAATVAAQRDTT 618
QY 301 VISHSPNTSYDTALEARIQKEEELIWMANKRCUDMEGRITKTHAQIIEKDAMIKVQQRS 360
DB 619 VISHSPNTSYDTALEARIQKEEELIWMANKRCUDMEGRITKTHAQIIEKDAMIKVQQRS 678
QY 361 RKEPSKTEQLSCMRPAKSLMSISNAGSGLLSHSTLTGSPIMBEKDDKSKWKSGLILG 420
DB 679 RKEPSKTEQLSCMRPAKSLMSISNAGSGLLSHSTLTGSPIMBEKDDKSKWKSGLILG 738
QY 421 GDYRAEYVPTSPVPSTPSTELLSAHKTGSRDCSTOTGRTSNTKTAAPVAPVAPVAA 480
DB 739 GDYRAEYVPTSPVPSTPSTELLSAHKTGSRDCSTOTGRTSNTKTAAPVAPVAPVAA 798
QY 481 AATAAAITATAA 492
DB 799 AATAAAITATAA 810
RESULT 4
ABP43965
ID ABP43965 standard; protein; 479 AA.
XX
XX ABP43965;
XX
XX 26-FEB-2003 (first entry)
XX Unidentified protein sequence.
XX
XX Neuroprotective; immunomodulator; cancer; chromosome X; cytostatic;
XX anti-inflammatory; gene therapy; nutritional supplement; wound; burn;

```

KW ulcer; Alzheimer's disease; Huntington's disease;
KW amyotrophic lateral sclerosis; autoimmune disorder; inflammation;
KW vulnervary.
OS Homo sapiens.
PN WO200231111-A2.
PD 18-APR-2002.
XX 11-OCT-2001; 2001WO-US027760.
XX 12-OCT-2000; 2000US-00687527.
XX (HYSE-) HYSEQ INC.
XX Tang YT, Liu C, Zhou P, Asundi V, Zhang J, Zhao QA, Ren F;
PI Xue AJ, Yang Y, Wehrman T, Drmanac RT;
XX
XX MPI; 2002-426278/45.
XX N-PSDB; ABQ61209.
XX New polypeptides and their encoded proteins, useful as nutritional
XX sources or supplements, or in gene therapy, particularly for treating
XX wounds, Alzheimer's disease, amyotrophic lateral sclerosis, cancer or
XX inflammation.
XX Claim 20; SEQ ID # 868; 357pp + Sequence Listing; English.
XX
XX The invention relates to 446 newly isolated polynucleotide sequences. The
XX activity of polynucleotides of the invention may be described as,
XX vulnerary, neuroprotective, immunomodulator, cytostatic and anti-
XX inflammatory. Compositions comprising nucleic acids of the invention are
XX useful for treating a mammalian subject, or as nutritional sources or
XX supplements. These are useful in gene therapy, particularly for treating
XX wounds, burns or ulcers, Alzheimer's disease, Huntington's disease,
XX amyotrophic lateral sclerosis, autoimmune disorders, cancer or
XX inflammation. The nucleic acids and polypeptides are also useful in
XX diagnostic and research methods. The sequences given in records ABP43544-
XX ABP43989 represent polypeptides encoded by polynucleotides of the
XX invention. NOTE: The sequence data for this patent did not form part of
XX the printed specification, but was obtained in electronic format directly
XX from WIPO at ftp.wipo.int/pub/published_pct_sequences
SQ Sequence 479 AA;

Query Match
Best Local Similarity 71.0%; Score 479; DB 5; Length 479;
Matches 479; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 197 MQQALVQLQAACEKREQLERLRLERLESLRIQQRGNCOPTNVSEYNAALMELLR 256
DB 1 MQQALVQLQAACEKREQLERLRLERLESLRIQQRGNCOPTNVSEYNAALMELLR 60
QY 257 EKEERILALEADMTKWEQKYLEENVMRHFALDAAATVAARDTTVISHSPNTSYDTALEA 316
DB 61 EKEERILALEADMTKWEQKYLEENVMRHFALDAAATVAARDTTVISHSPNTSYDTALEA 120
QY 317 RIQKEEERILMANKECLDMEGRITLHAQIKLQKAMIKVLQQRSRKPSKTEQLSCMRPA 376
DB 121 RIQKEEERILMANKECLDMEGRITLHAQIKLQKAMIKVLQQRSRKPSKTEQLSCMRPA 180
QY 377 KSLMISINAGSGLLSHSSTLTGSPIMEKRDKSKWKSGLILGCDYRBYVSTSPVP 436
DB 181 KSLMISINAGSGLLSHSSTLTGSPIMEKRDKSKWKSGLILGCDYRBYVSTSPVP 240
QY 437 PSTPLLSAHSKGTGRDSCSTQTERGTESNKTAAVAPISVPAPVAAAATAATATATITTT 496
DB 241 PSTPLLSAHSKGTGRDSCSTQTERGTESNKTAAVAPISVPAPVAAAATAATATATITTT 300
QY 497 TWVAAAPVAVAAAAAPAAAAAPSPTAAATAAATAAVSPAAAGQIPAAASVAAAAVAPSA 556
DB 301 TWVAAAPVAVAAAAAPAAAAAPSPTAAATAAATAAVSPAAAGQIPAAASVAAAAVAPSA 360

QY 557 AAAYQVAPAPAPVPAPALVPVPAPAAAQASAPAQOAPTSAPAVAPPTAPTTPPAVQA 616
DB 361 AAAYQVAPAPAPVPAPALVPVPAPAAAQASAPAQOAPTSAPAVAPPTAPTTPPAVQA 420
QY 617 EVPASPATGPGPHRLSPSLTCNPKDGTGPFVFSNLTERTKPIQILQGBDAEWVEYLI 675
DB 421 EVPASPATGPGPHRLSPSLTCNPKDGTGPFVFSNLTERTKPIQILQGBDAEWVEYLI 479

RESULT 5
ABG17959
ID ABG17959 standard; protein; 759 AA.
XX
XX ABG17959;
XX 18-FEB-2002 (first entry)
XX Novel human diagnostic protein #17950.
XX Human; chromosome mapping; gene mapping; gene therapy; forensic;
XX food supplement; medical imaging; diagnostic; genetic disorder.
XX Homo sapiens.
XX WO200175067-A2.
XX
XX 11-OCT-2001.
XX 30-MAR-2001; 2001WO-US008631.
XX 31-MAR-2000; 2000US-00540217.
XX 23-AUG-2000; 2000US-00649167.
XX (HYSE-) HYSEQ INC.
XX Drmanac RT, Liu C, Tang YT;
XX WPI; 2001-639362/73.
XX N-PSDB; AAS82146.
XX New isolated polynucleotide and encoded polypeptides, useful in
XX diagnostics, forensics, gene mapping, identification of mutations
XX responsible for genetic disorders or other traits and to assess
XX biodiversity.
XX Claim 20; SEQ ID NO 48318; 103pp; English.
XX
XX The invention relates to isolated polynucleotide (I) and polypeptide (II)
XX sequences. (I) is useful as hybridisation probes, polymerase chain
XX reaction (PCR) primers, oligomers, and for chromosome and gene mapping,
XX and in recombinant production of (II). The polynucleotides are also used
XX in diagnostics as expressed sequence tags for identifying expressed
XX genes. (I) is useful in gene therapy techniques to restore normal
XX activity of (II) or to treat disease states involving (II). (II) is
XX useful for generating antibodies against it, detecting or quantitating a
XX polypeptide in tissue, as molecular weight markers and as a food
XX supplement. (II) and its binding partners are useful in medical imaging
XX of sites expressing (II). (I) and (II) are useful for treating disorders
XX involving aberrant protein expression or biological activity. The
XX polypeptide and polynucleotide sequences have applications in
XX diagnostics, forensics, gene mapping, identification of mutations
XX responsible for genetic disorders or other traits to assess biodiversity
XX and to produce other types of data and products dependent on DNA and
XX amino acid sequences. ABG30010-ABG30377 represent novel human diagnostic
XX amino acid sequences of the invention. Note: The sequence data for this
XX patent did not appear in the printed specification, but was obtained in
XX electronic format directly from WIPO at
XX ftp.wipo.int/pub/published_pct_sequences
XX Sequence 759 AA;

Query Match 27.1%; Score 183; DB 4; Length 759;

Best Local Similarity 100.0%; Pred. No. 8.5e-157;
Matches 183; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MPRAQSSASQVPADPFAIVSRAQWVEILSDENRLRQEGCGYKVARLQKVETEI 60
DB 525 MPRAQSSASQVPADPFAIVSRAQWVEILSDENRLRQEGCGYKVARLQKVETEI 584
QY 61 QRVSEAYENLVKSSKRALEKAVNKLGEIRWHDFFNDRDLERLETANKOLAEKEYEG 120
DB 585 QRVSEAYENLVKSSKRALEKAVNKLGEIRWHDFFNDRDLERLETANKOLAEKEYEG 644
QY 121 SEDTRKTIQSILFAKNKESQREKEKLEAELEATARGTNEQRRHIEIRQALSNAQAKVVKL 180
DB 645 SEDTRKTIQSILFAKNKESQREKEKLEAELEATARGTNEQRRHIEIRQALSNAQAKVVKL 704
QY 181 BEE 183
DB 705 BEE 707

RESULT 6
AAY54054
ID AAY54054 standard; protein; 143 AA.
AC AAY54054;
DT 27-MAR-2000 (first entry)
CX Angiostatin-binding domain of ABP-1, designated Big-3.
KW Human; angiogenesis-associated protein; plasminogen; ABP-1;
KW kringle domain; angiotensin; plasminogen receptor;
KW angiogenesis-related disease; tumor; diabetes; rheumatoid arthritis;
KW inflammatory disease; psoriasis; chronic inflammation; intestine; asthma;
KW obesity; gene therapy; angiotensin-binding domain; Big-3.
DS Homo sapiens.
CX WO9966038-A1.
CX 23-DEC-1999.
CX 11-JUN-1999; 98WO-EP004109.
CX 15-JUN-1998; 98SE-00002130.
CX 15-JUN-1998; 98US-0089266P.
CX 17-DEC-1998; 98SE-00004372.
CX 29-DEC-1998; 98US-0114386P.
CX (PHAA) PHARMACIA & UPJOHN AB.
CX Holmgren L, Troyanovsky B;
CX WPI; 2000-106099/09.
CX Novel human protein useful for treating angiogenesis associated diseases
CX or disorders.
CX Claim 6; Page 49; 58pp; English.
CX The present sequence represents the angiotensin-binding domain,
CX designated Big-3, of a human angiogenesis-associated protein. The protein
CX binds an N-terminal fragment of plasminogen. The ABP-1 protein is
CX designated ABP-1, and binds the first 4 kringle domains (K1-K4) and/or
CX kringle 5 (K5) of plasminogen. These four kringle domains comprise
CX angiotensin. The ABP-1 protein acts as a receptor for plasminogen. A
CX polymorphic variant of ABP-1 is also described, in AAY54053. ABP-1 can be
CX used to manufacture medicaments for treating angiogenesis-related
CX diseases or disorders, such as tumor conditions, diabetes, rheumatoid
CX arthritis, and even some inflammatory diseases such as psoriasis, chronic
CX inflammation of the intestine, asthma, etc. . The protein may also be able
CX to treat and cure, or prevent, obesity. The ABP-1 DNA can be used in gene
CX therapy techniques

XX Sequence 143 AA;
SQ Query Match 21.2%; Score 143; DB 3; Length 143;
Best Local Similarity 100.0%; Pred. No. 3.7e-121;
Matches 143; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 462 ESNKTAAVAPISVPAPVAAATAAATATATATATTTTWWAAAPVAVAAAPAAAAAASPA 521
DB 1 ESNKTAAVAPISVPAPVAAATAAATATATATATTTTWWAAAPVAVAAAPAAAAAASPA 60
QY 522 TAAATAAAVSPAAAGQIPAAASVASAAVAPSAASAAAQVAPAPVAPALVPPVAP 581
DB 61 TAAATAAAVSPAAAGQIPAAASVASAAVAPSAASAAAQVAPAPVAPALVPPVAP 120
QY 582 AAAQASAPAQTAQTSAPAVAPT 604
DB 121 AAAQASAPAQTAQTSAPAVAPT 143

RESULT 7
AAM16746
ID AAM16746 standard; protein; 208 AA.
XX AAM16746;
XX AC AAM16746;
DT 12-OCT-2001 (first entry)
CX Peptide #3180 encoded by probe for measuring cervical gene expression.
KW Probe; human; microarray; gene expression; cervical epithelial cell;
KW cervical cancer.
OS Homo sapiens.
CX WO200157278-A2.
CX 09-AUG-2001.
CX 30-JAN-2001; 2001WO-US000670.
CX 04-FEB-2000; 2000US-0180312P.
CX 26-MAY-2000; 2000US-0207456P.
CX 30-JUN-2000; 2000US-00808408.
CX 03-AUG-2000; 2000US-00632366.
CX 21-SEP-2000; 2000US-0234687P.
CX 27-SEP-2000; 2000US-0236359P.
CX 04-OCT-2000; 2000GB-00024263.
CX (MOLE-) MOLECULAR DYNAMICS INC.
CX Penn SG, Hanzel DK, Chen W, Rank DR;
CX WPI; 2001-488901/53.
CX Human genome-derived single exon nucleic acid probes useful for analyzing
CX gene expression in human cervical epithelial cells.
CX Claim 27; SEQ ID NO 21572; 487pp; English.
CX The present invention relates to human single exon nucleic acid probes
CX (SENP: see AAI10068-AA128459). The present sequence is a peptide encoded
CX by one such probe. The SENPs are derived from human HeLa cells. The SENPs
CX can be used to produce a single exon microarray, which can be used for
CX measuring human gene expression in a sample derived from human cervical
CX epithelial cells. By measuring gene expression, the probes are therefore
CX useful in grading and/or staging of diseases of the cervix, notably
CX cervical cancer. Note: The sequence data for this patent did not form
CX part of the printed specification, but was obtained in electronic format
CX directly from WIPO at ftp.wipo.int/pub/published_pct_sequences
CX Sequence 208 AA;
SQ

```
Query Match      8.1%; Score 55; DB 4; Length 208;
Best Local Similarity 100.0%; Pred. No. 3.4e-41;
Matches 55; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 MPRAQPSASIQVPADPPAIVSRAQQWVEILSDENRNLRQELGECYKVARLOK 55
Db      154 MPRAQPSASIQVPADPPAIVSRAQQWVEILSDENRNLRQELGECYKVARLOK 208

RESULT 8
ABB35732
ID ABB35732 standard; peptide; 208 AA.
XX
AC ABB35732;
XX
DT 04-FEB-2002 (first entry)
XX
DE Peptide #3238 encoded by human foetal liver single exon probe.
XX
KW Human; foetal liver; gene expression; single exon nucleic acid probe.
XX
OS Homo sapiens.
XX
PN WO200157277-A2.
XX
PD 09-AUG-2001.
XX
PF 30-JAN-2001; 2001WO-US000669.
XX
PR 04-FEB-2000; 2000US-0180312P.
PR 26-MAY-2000; 2000US-0207456P.
PR 30-JUN-2000; 2000US-0060840B.
PR 03-AUG-2000; 2000US-00632366.
PR 21-SEP-2000; 2000US-0234687P.
PR 27-SEP-2000; 2000US-0236359P.
PR 04-OCT-2000; 2000GB-00024263.
XX
PA (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX
PD WPI; 2001-483447/52.
XX
DR Human genome-derived single exon nucleic acid probes useful for analyzing
PT gene expression in human foetal liver.
XX
PS Claim 27; SEQ ID NO 28367; 639pp + Sequence Listing; English.
XX
CC The invention relates to a single exon nucleic acid probe for measuring
CC human gene expression in a sample derived from human foetal liver. The
CC single exon nucleic acid probes may be used for predicting, measuring and
CC displaying gene expression in samples derived from human foetal liver. The
CC present sequence is a peptide encoded by a single exon nucleic acid probe
CC of the invention. Note: The sequence data for this patent did not form
CC part of the printed specification, but was obtained in electronic format
CC directly from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 208 AA;

Query Match      8.1%; Score 55; DB 4; Length 208;
Best Local Similarity 100.0%; Pred. No. 3.4e-41;
Matches 55; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 MPRAQPSASIQVPADPPAIVSRAQQWVEILSDENRNLRQELGECYKVARLOK 55
Db      154 MPRAQPSASIQVPADPPAIVSRAQQWVEILSDENRNLRQELGECYKVARLOK 208

RESULT 9
AAM29234
ID AAM29234 standard; protein; 208 AA.
XX
AC AAM29234;
```

```
XX
DT 17-OCT-2001 (first entry)
XX
DE Peptide #3271 encoded by probe for measuring placental gene expression.
XX
KW Probe; microarray; human; placenta; antenatal diagnosis;
XX
OS genetic disorder.
XX
OS Homo sapiens.
XX
PN WO200157272-A2.
XX
PD 09-AUG-2001.
XX
PF 30-JAN-2001; 2001WO-US000663.
XX
PR 04-FEB-2000; 2000US-0180312P.
PR 26-MAY-2000; 2000US-0207456P.
PR 30-JUN-2000; 2000US-0060840B.
PR 03-AUG-2000; 2000US-00632366.
PR 21-SEP-2000; 2000US-0234687P.
PR 27-SEP-2000; 2000US-0236359P.
PR 04-OCT-2000; 2000GB-00024263.
XX
PA (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX
PD WPI; 2001-488897/53.
XX
DR Human genome-derived single exon nucleic acid probes useful for analyzing
PT gene expression in human placenta.
XX
PS Claim 27; SEQ ID NO 29503; 654pp; English.
XX
CC The present invention relates to single exon nucleic acid probes (SENP:
CC see A131315-A137546). The present sequence is a peptide encoded by one
CC such probe. The probes are useful for producing a microarray for
CC predicting, measuring and displaying gene expression in samples derived
CC from human placenta. The probes are useful for antenatal diagnosis of
CC human genetic disorders
XX
SQ Sequence 208 AA;

Query Match      8.1%; Score 55; DB 4; Length 208;
Best Local Similarity 100.0%; Pred. No. 3.4e-41;
Matches 55; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 MPRAQPSASIQVPADPPAIVSRAQQWVEILSDENRNLRQELGECYKVARLOK 55
Db      154 MPRAQPSASIQVPADPPAIVSRAQQWVEILSDENRNLRQELGECYKVARLOK 208

RESULT 10
ABB30567
ID ABB30567 standard; peptide; 208 AA.
XX
AC ABB30567;
XX
DT 01-FEB-2002 (first entry)
XX
DE Peptide #3218 encoded by breast cell single exon nucleic acid probe.
XX
KW Human; microarray; single exon probe; gene expression; breast; disease;
XX
OS cancer.
XX
OS Homo sapiens.
XX
PN WO200157271-A2.
XX
PD 09-AUG-2001.
XX
PF 30-JAN-2001; 2001WO-US000662.
```

X 04-FEB-2000; 2000US-0180312P.
 R 26-MAY-2000; 2000US-0207456P.
 R 30-JUN-2000; 2000US-00608408.
 R 03-AUG-2000; 2000US-00632366.
 R 21-SEP-2000; 2000US-0234587P.
 R 27-SEP-2000; 2000US-0236359P.
 R 04-OCT-2000; 2000GB-00024263.
 X (MOLE-) MOLECULAR DYNAMICS INC.
 X Penn SG, Hanzel DK, Chen W, Rank DR;
 X WPI; 2001-496933/54.
 X
 T New spatially-addressable set of single exon nucleic acid probes, useful
 T for measuring gene expression in sample derived from human breast,
 T comprises number of single exon nucleic acid probes.
 X
 S Claim 27; SEQ ID NO 13535; 327pp + Sequence Listing; English.
 X
 C The invention relates to a spatially-addressable set of single exon
 C nucleic acid probes for measuring gene expression in a sample derived
 C from human breast and Br 474 cells. The method involves contacting the
 C probes with a collection of detectably labelled nucleic acids derived
 C from mRNA of human breast, and then measuring the label bound to each
 C probe of the microarray. The probes are useful for verifying the
 C expression of regions of genomic DNA predicted to encode proteins. They
 C are useful for gene discovery, and for determining predisposition and/or
 C prognosing breast disease. Gene expression analysis is useful for
 C assessing the toxicity of chemical agents on cells. The microarray of
 C this invention presents a far greater diversity of probes for measuring
 C gene expression, with far less bias than expressed sequence tag
 C microarrays. The method is suitable for rapid production of functional
 C information from genomic sequence. The present sequence is a peptide
 C encoded by a single exon nucleic acid probe of the invention. Note: The
 C sequence data for this patent did not form part of the printed
 C specification, but was obtained in electronic format directly from WIPO
 C at ftp.wipo.int/pub/published_pct_sequences
 X
 Q Sequence 208 AA;
 Query Match 8.1%; Score 55; DB 4; Length 208;
 Best Local Similarity 100.0%; Pred. No. 3.4e-41;
 Matches 55; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Y 1 MPRAQSSASVQVPADPPFAIVSRAQQMVILLSDENRNLQELGCTYKVARLQK 55
 b 154 MPRAQSSASVQVPADPPFAIVSRAQQMVILLSDENRNLQELGCTYKVARLQK 208
 RESULT 11
 B21158
 D ABB21158 standard; protein; 208 AA.
 X
 C ABB21158;
 X
 T 23-JAN-2002 (first entry)
 X
 DE Protein #3157 encoded by probe for measuring heart cell gene expression.
 X
 W Human; gene expression; heart; microarray; vascular system;
 W cardiovascular disease; hypertension; cardiac arrhythmia;
 W congenital heart disease.
 X
 NS Homo sapiens.
 X
 X WO200157274-A2.
 X
 X 09-AUG-2001.
 X
 X 30-JAN-2001; 2001WO-US000666.
 X
 X (MOLE-) MOLECULAR DYNAMICS INC.

PR 04-FEB-2000; 2000US-0180312P.
 PR 26-MAY-2000; 2000US-0207456P.
 PR 30-JUN-2000; 2000US-00608408.
 PR 03-AUG-2000; 2000US-00632366.
 PR 21-SEP-2000; 2000US-0234587P.
 PR 27-SEP-2000; 2000US-0236359P.
 PR 04-OCT-2000; 2000GB-00024263.
 X (MOLE-) MOLECULAR DYNAMICS INC.
 X Penn SG, Hanzel DK, Chen W, Rank DR;
 X WPI; 2001-498899/53.
 X
 T Single exon nucleic acid probes for analyzing gene expression in human
 T hearts.
 X
 PS Claim 15; SEQ ID NO 22928; 530pp; English.
 X
 C The present invention relates to single exon nucleic acid probes for
 C measuring human gene expression in a sample derived from human heart (see
 C AB21535-AB41305). The present sequence is a protein encoded by one such
 C probe. The probes may be used for predicting, measuring and displaying
 C gene expression in samples derived from the human heart via microarrays.
 C By measuring gene expression, the probes are useful for predicting,
 C diagnosing, grading, staging, monitoring and prognosing diseases of the
 C human heart and vascular system e.g. cardiovascular disease,
 C hypertension, cardiac arrhythmias and congenital heart disease. Note: The
 C sequence data for this patent did not form part of the printed
 C specification, but was obtained in electronic format directly from WIPO
 C at ftp.wipo.int/pub/published_pct_sequences
 X
 SQ Sequence 208 AA;
 Query Match 8.1%; Score 55; DB 4; Length 208;
 Best Local Similarity 100.0%; Pred. No. 3.4e-41;
 Matches 55; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Q 1 MPRAQSSASVQVPADPPFAIVSRAQQMVILLSDENRNLQELGCTYKVARLQK 55
 b 154 MPRAQSSASVQVPADPPFAIVSRAQQMVILLSDENRNLQELGCTYKVARLQK 208
 RESULT 12
 AAM68923
 ID AAM68923 standard; protein; 208 AA.
 X
 AC AAM68923;
 X
 DT 06-NOV-2001 (first entry)
 X
 DE Human bone marrow expressed probe encoded protein SEQ ID NO: 29229.
 X
 KW Human; bone marrow expressed exon; gene expression analysis; probe;
 KW microarray; cancer; leukaemia; lymphoma; myeloma.
 X
 OS Homo sapiens.
 X
 PN WO200157276-A2.
 X
 PD 09-AUG-2001.
 X
 X 30-JAN-2001; 2001WO-US000668.
 X
 PR 04-FEB-2000; 2000US-0180312P.
 PR 26-MAY-2000; 2000US-0207456P.
 PR 30-JUN-2000; 2000US-00608408.
 PR 03-AUG-2000; 2000US-00632366.
 PR 21-SEP-2000; 2000US-0234587P.
 PR 27-SEP-2000; 2000US-0236359P.
 PR 04-OCT-2000; 2000GB-00024263.
 X (MOLE-) MOLECULAR DYNAMICS INC.

XX Penn SG, Hanzel DX, Chen W, Rank DR;
 XX WPI; 2001-488900/53.
 XX Human genome-derived single exon nucleic acid probes useful for analyzing
 PT gene expression in human bone marrow.
 XX Example 4; SEQ ID NO 29229; 658pp + Sequence Listing; English.
 XX The present invention provides a number of single exon nucleic acid
 CC probes which are derived from genomic sequences expressed in the human
 CC bone marrow. They can be used to measure gene expression in bone marrow
 CC samples, which may enable the improved diagnosis and treatment of cancers
 CC such as lymphoma, leukaemia and myeloma. The present sequence is a
 CC protein encoded by one of the probes of the invention
 XX Sequence 208 AA;
 SQ

Query Match 8.1%; Score 55; DB 4; Length 208;
 Best Local Similarity 100.0%; Pred. NO. 3.4e-41;
 Matches 55; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 XX

1 MRAQPSASQVPADPFAIVSRAQQWVILSDENLNRLQELGGCYKVARLQK 55
 154 MRAQPSASQVPADPFAIVSRAQQWVILSDENLNRLQELGGCYKVARLQK 208

RESULT 13
 AAM56541
 ID AAM56541 standard; protein; 208 AA.
 AC AAM56541;
 XX

05-NOV-2001 (first entry)
 XX Human brain expressed single exon probe encoded protein SEQ ID NO: 28646.
 XX Human; brain expressed exon; gene expression analysis; probe; microarray;
 KW Alzheimer's disease; multiple sclerosis; schizophrenia; epilepsy; cancer.
 XX Homo sapiens.
 XX WO200157275-A2.
 XX 09-AUG-2001.
 XX 30-JAN-2001; 2001WO-US000667.
 XX 04-FEB-2000; 2000US-0180312P.
 PR 26-MAY-2000; 2000US-0207456P.
 PR 30-JUN-2000; 2000US-00608408.
 PR 03-AUG-2000; 2000US-00632366.
 PR 21-SEP-2000; 2000US-0234687P.
 PR 27-SEP-2000; 2000US-0234359P.
 PR 04-OCT-2000; 2000GB-00024253.
 XX (MOLE-) MOLECULAR DYNAMICS INC.
 PA Penn SG, Hanzel DX, Chen W, Rank DR;
 PI WPI; 2001-483446/52.
 XX Single exon nucleic acid probes for analyzing gene expression in human
 PT brains.
 XX Example 4; SEQ ID NO 28646; 650pp + Sequence Listing; English.
 XX The present invention provides a number of single exon nucleic acid
 CC probes which are derived from genomic sequences expressed in the human
 CC brain. They can be used to measure gene expression in brain cell samples,
 CC which may enable the diagnosis and improved treatment of nervous system
 CC diseases such as Alzheimer's disease, multiple sclerosis, schizophrenia,
 CC

CC epilepsy and cancers. The present sequence is a protein encoded by one of
 CC the probes of the invention
 XX Sequence 208 AA;
 SQ

Query Match 8.1%; Score 55; DB 4; Length 208;
 Best Local Similarity 100.0%; Pred. NO. 3.4e-41;
 Matches 55; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 XX

1 MRAQPSASQVPADPFAIVSRAQQWVILSDENLNRLQELGGCYKVARLQK 55
 154 MRAQPSASQVPADPFAIVSRAQQWVILSDENLNRLQELGGCYKVARLQK 208

RESULT 14
 ABG50588
 ID ABG50588 standard; peptide; 208 AA.
 XX ABG50588;
 XX

25-FEB-2003 (first entry)
 XX Human liver peptide, SEQ ID NO 29236.
 XX Human; liver; cirrhosis; hyperlipoproteinaemia; hyperlipidaemia;
 KW hypercholesterolaemia; coronary heart disease.
 XX Homo sapiens.
 XX WO200157273-A2.
 XX 09-AUG-2001.
 XX 30-JAN-2001; 2001WO-US000664.
 XX 04-FEB-2000; 2000US-0180312P.
 PR 26-MAY-2000; 2000US-0207456P.
 PR 30-JUN-2000; 2000US-00608408.
 PR 03-AUG-2000; 2000US-00632366.
 PR 21-SEP-2000; 2000US-0234687P.
 PR 27-SEP-2000; 2000US-0234359P.
 PR 04-OCT-2000; 2000GB-00024253.
 XX (MOLE-) MOLECULAR DYNAMICS INC.
 PA Penn SG, Hanzel DK, Chen W, Rank DR;
 PI WPI; 2001-488898/53.
 XX Human genome-derived single exon nucleic acid probes useful for analyzing
 PT gene expression in human adult liver.
 XX Claim 27; SEQ ID NO 29236; 658pp; English.
 XX The invention relates to a single exon nucleic acid probe (SENP) (I) for
 CC measuring human gene expression in a sample derived from human adult
 CC liver, comprising one of 13109 defined nucleotide sequences given in the
 CC specification (or complements/ fragments). The probe hybridises at high
 CC stringency to a nucleic acid molecule expressed in the human adult liver.
 CC (I) may be used for predicting, measuring and displaying gene expression
 CC in samples derived from human adult liver. The genes identified may be
 CC involved in genetic liver diseases such as cirrhosis,
 CC hyperlipoproteinaemia, hyperlipidaemia and hypercholesterolaemia which is
 CC associated with coronary heart disease. ABG47148-ABG59930 represent human
 CC liver single exon encoded peptides of the invention. Note: The sequence
 CC information for this patent does not appear in the printed specification
 CC but was obtained in electronic format directly from WIPO at
 CC ftp.wipo.int/pub/published_pct_sequences
 XX Sequence 208 AA;
 SQ

Query Match 8.1%; Score 55; DB 4; Length 208;
 Best Local Similarity 100.0%; Pred. NO. 3.4e-41;
 XX

Job time : 63 secs

Matches 55; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 MPRAQSSASYQVPADPPAIVSRAQQMVEILSDENRNLRQELGCGYKVARLQK 55
154 MPRAQSSASYQVPADPPAIVSRAQQMVEILSDENRNLRQELGCGYKVARLQK 208

RESULT 15

AM04463
D AA004463 standard; protein; 208 AA.

X C

X C

X C

X C

X C

X C

X C

X C

X C

X C

X C

X C

X C

X C

X C

X C

X C

X C

X C

X C

X C

X C

X C

X C

X C

X C

X C

X C

X C

X C

X C

X C

X C

X C

09-OCT-2001 (first entry)
Peptide #3145 encoded by probe for measuring breast gene expression.

Probe; human; breast disease; breast cancer; development disorder;
inflammatory disease; proliferative breast disease; non-carcinoma tumour.

Homo sapiens.

US

WO200157270-A2.

X C

X C

X C

X C

X C

X C

X C

X C

X C

X C

X C

X C

X C

X C

X C

X C

X C

X C

X C

X C

X C

X C

X C

X C

X C

X C

X C

29-JAN-2001; 2001WO-US000661.

04-FEB-2000; 2000US-0180312P.

28-MAY-2000; 2000US-0207456P.

30-JUN-2000; 2000US-00608408.

03-AUG-2000; 2000US-00632366.

21-SEP-2000; 2000US-0234687P.

27-SEP-2000; 2000US-0236359P.

04-OCT-2000; 2000GB-00024263.

(MOLE-) MOLECULAR DYNAMICS INC.

Penn SG, Hanzel DK, Chen W, Rank DR;

WPI; 2001-476286/51.

Novel single exon nucleic acid probe used to measuring gene expression in a human breast.

Claim 27; SEQ ID NO 13203; 322pp; English.

The present invention relates to novel single exon nucleic acid probes (see AA10010-AA110067). The present sequence is a peptide encoded by one such probe. The probes are useful for measuring human gene expression in a human breast sample, where the probe hybridises at high stringency to a nucleic acid expressed in the human breast. The probes are useful for predicting, diagnosing, grading, staging, monitoring and prognosing diseases of the human breast, particularly those diseases with polygenic aetiology. The diseases include: breast cancer; disorders of development, inflammatory diseases of the breast, fibrocystic changes, proliferative breast disease and non-carcinoma tumours. Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published_pct_sequences

Sequence 208 AA;

Query Match 8.1%; Score 55; DB 4; Length 208;
Best Local Similarity 100.0%; Pred.No. 3.4e-41;
Matches 55; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 MPRAQSSASYQVPADPPAIVSRAQQMVEILSDENRNLRQELGCGYKVARLQK 55
154 MPRAQSSASYQVPADPPAIVSRAQQMVEILSDENRNLRQELGCGYKVARLQK 208

Search completed: June 16, 2004, 19:25:42